```
                        GenCore version 5.1.6
                Copyright (c) 1993 - 2004  Compugen Ltd.

OM protein - protein search, using sw model

Run on:         September 27, 2004, 19:19:42 ; Search time 120 Seconds
                                               (without alignments)
                                               1254.185 Million cell updates/sec

Title:          US-09-886-954A-1
Perfect score:  2457
Sequence:       1 MTPEDPEETQPLLGPPGGSA..........CVPETKGKTLEQITAHFEGR 477

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters:      1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       SPTREMBL_25:*
                1:  sp_archea:*
                2:  sp_bacteria:*
                3:  sp_fungi:*
                4:  sp_human:*
                5:  sp_invertebrate:*
                6:  sp_mammal:*
                7:  sp_mhc:*
                8:  sp_organelle:*
                9:  sp_phage:*
                10:  sp_plant:*
                11:  sp_rodent:*
                12:  sp_virus:*
                13:  sp_vertebrate:*
                14:  sp_unclassified:*
                15:  sp_rvirus:*
                16:  sp_bacteriap:*
                17:  sp_archeap:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                               SUMMARIES
                  %
Result          Query
  No.   Score   Match Length DB  ID                  Description
 -------------------------------------------------------------------------
    1    2457   100.0    477  4  Q8WUZ9               Q8wuz9 homo sapien
    2  1411.5    57.4    498 13   Q7ZWH3               Q7zwh3 brachydanio
    3    1407    57.3    482 13   Q8AYP6               Q8ayp6 gallus gall
    4    1167    47.5    266 11   Q8VDJ4               Q8vdj4 mus musculu
    5    1131    46.0    248  4  Q8WZ05               Q8wz05 homo sapien
    6  1125.5    45.8    246  6  Q8SPS1               Q8sps1 ovis aries
    7     946    38.5    497 11   Q8BTN2               Q8btn2 mus musculu
    8     894    36.4    445  4  Q8NCC2               Q8ncc2 homo sapien
    9   804.5    32.7    488  5  Q8MKK4               Q8mkk4 drosophila
   10   789.5    32.1    506  5  Q86P59               Q86p59 drosophila
   11   788.5    32.1    489  5  Q9V609               Q9v609 drosophila
   12   788.5    32.1    857  5  Q9V608               Q9v608 drosophila
   13     788    32.1    539  5  Q9VU17               Q9vu17 drosophila
   14     787    32.0    471  5  Q8IQH6               Q8iqh6 drosophila
   15   764.5    31.1    433  5  Q9V610               Q9v610 drosophila
   16     736    30.0    488 10   Q93YP9               Q93yp9 arabidopsis
   17     734    29.9    487 10   Q9FRL3               Q9frl3 arabidopsis
   18     727    29.6    465  5  Q9V848               Q9v848 drosophila
   19   724.5    29.5    515 10   Q9LN48               Q9ln48 arabidopsis
   20     723    29.4    501 10   Q9XHW3               Q9xhw3 oryza sativ
   21     722    29.4    490 10   Q39416               Q39416 beta vulgar
   22   710.5    28.9    496  5  Q8T0T6               Q8t0t6 drosophila
   23     684    27.8    482 10   Q8VZI5               Q8vzi5 arabidopsis
   24     675    27.5    482 10   Q8LBI9               Q8lbi9 arabidopsis
   25     670    27.3    463 10   Q8LFR8               Q8lfr8 arabidopsis
   26   666.5    27.1    444  5  Q9V3G0               Q9v3g0 drosophila
   27     662    26.9    463 10   Q9ZU87               Q9zu87 arabidopsis
   28   653.5    26.6    463 10   Q9MAA4               Q9maa4 arabidopsis
   29   650.5    26.5    480 10   Q9LEG2               Q9leg2 lycopersico
   30     647    26.3    491  5  Q9VNJ2               Q9vnj2 drosophila
   31     607    24.7    471 16   Q8Z401               Q8z401 salmonella
   32     607    24.7    472 16   Q8ZMA0               Q8zma0 salmonella
   33   604.5    24.6    596 10   Q7XIZ0               Q7xiz0 oryza sativ
   34     603    24.5    472 16   Q7UBR3               Q7ubr3 shigella fl
   35     603    24.5    507 16   Q83QB4               Q83qb4 shigella fl
   36     601    24.5    515 16   Q8FE99               Q8fe99 escherichia
   37   594.5    24.2    460 16   Q89ZI0               Q89zi0 bacteroides
   38   593.5    24.2    409 16   Q83EH4               Q83eh4 coxiella bu
   39   587.5    23.9    451 16   Q7UBN5               Q7ubn5 shigella fl
   40   587.5    23.9    464 16   Q8XCW7               Q8xcw7 escherichia
   41   587.5    23.9    468 16   Q83Q92               Q83q92 shigella fl
   42     585    23.8    509 10   Q8VZR6               Q8vzr6 arabidopsis
   43   584.5    23.8    496 10   O65799               O65799 arabidopsis
   44   583.5    23.7    464 16   Q8XEV6               Q8xev6 salmonella
   45     582    23.7    521 10   O22848               O22848 arabidopsis

                              ALIGNMENTS


RESULT 1
Q8WUZ9
ID   Q8WUZ9      PRELIMINARY;      PRT;   477 AA.
AC   Q8WUZ9;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Skin;
RA   Strausberg R.;
RL   Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
DR   EMBL; BC019043; AAH19043.1; -.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0005351; F:sugar porter activity; IEA.
DR   GO; GO:0005215; F:transporter activity; IEA.
DR   GO; GO:0008643; P:carbohydrate transport; IEA.
DR   InterPro; IPR007114; MFS.
DR   InterPro; IPR005828; Sub_transporter.
DR   InterPro; IPR003663; Sugar_transpt.
DR   InterPro; IPR005829; Sug_transporter.
DR   Pfam; PF00083; sugar_tr; 1.
DR   PRINTS; PR00171; SUGRTRNSPORT.
DR   TIGRFAMs; TIGR00879; SP; 1.
DR   PROSITE; PS50850; MFS; 1.
DR   PROSITE; PS00216; SUGAR_TRANSPORT_1; 2.
DR   PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.
KW   Hypothetical protein; Sugar transport; Transmembrane; Transport.
SQ   SEQUENCE   477 AA;  50819 MW;  0B480F94AF063316 CRC64;

  Query Match            100.0%;  Score 2457;  DB 4;  Length 477;
  Best Local Similarity  100.0%;  Pred. No. 2.4e-157;
```

```
 Matches  477;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MTPEDPEETQPLLGPPGGSAPRGRRVFLAAFAAALGPLSFGFALGYSSPAIPSLQRAAPP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MTPEDPEETQPLLGPPGGSAPRGRRVFLAAFAAALGPLSFGFALGYSSPAIPSLQRAAPP 60

Qy         61 APRLDDAAASWFGAVVTLGAAAGGVLGGWLVDRAGRKLSLLLCSVPFVAGFAVITAAQDV 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 APRLDDAAASWFGAVVTLGAAAGGVLGGWLVDRAGRKLSLLLCSVPFVAGFAVITAAQDV 120

Qy        121 WMLLGGRLLTGLACGVASLVAPVYISEIAYPAVRGLLGSCVQLMVVVGILLAYLAGWVLE 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 WMLLGGRLLTGLACGVASLVAPVYISEIAYPAVRGLLGSCVQLMVVVGILLAYLAGWVLE 180

Qy        181 WRWLAVLGCVPPSLMLLLMCFMPETPRFLLTQHRRQEAMAALRFLWGSEQGWEDPPIGAE 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 WRWLAVLGCVPPSLMLLLMCFMPETPRFLLTQHRRQEAMAALRFLWGSEQGWEDPPIGAE 240

Qy        241 QSFHLALLRQPGIYKPFIIGVSLMAFQQLSGVNAVMFYAETIFEEAKFKDSSLASVVVGV 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QSFHLALLRQPGIYKPFIIGVSLMAFQQLSGVNAVMFYAETIFEEAKFKDSSLASVVVGV 300

Qy        301 IQVLFTAVAALIMDRAGRRLLLVLSGVVMVFSTSAFGAYFKLTQGGPGNSSHVAISAPVS 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 IQVLFTAVAALIMDRAGRRLLLVLSGVVMVFSTSAFGAYFKLTQGGPGNSSHVAISAPVS 360

Qy        361 AQPVDASVGLAWLAVGSMCLFIAGFAVGWGPIPWLLMSEIFPLHVKGVATGICVLTNWLM 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AQPVDASVGLAWLAVGSMCLFIAGFAVGWGPIPWLLMSEIFPLHVKGVATGICVLTNWLM 420

Qy        421 AFLVTKEFSSLMEVLRPYGAFWLASAFCIFSVLFTLFCVPETKGKTLEQITAHFEGR 477
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AFLVTKEFSSLMEVLRPYGAFWLASAFCIFSVLFTLFCVPETKGKTLEQITAHFEGR 477

RESULT 2
Q7ZWH3
ID   Q7ZWH3      PRELIMINARY;      PRT;   498 AA.
AC   Q7ZWH3;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein.
OS   Brachydanio rerio (Zebrafish) (Danio rerio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Body;
RA   Strausberg R.;
RL   Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC049409; AAH49409.1; -.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0005351; F:sugar porter activity; IEA.
DR   GO; GO:0005215; F:transporter activity; IEA.
DR   GO; GO:0008643; P:carbohydrate transport; IEA.
DR   InterPro; IPR007114; MFS.
DR   InterPro; IPR005828; Sub_transporter.
DR   InterPro; IPR003663; Sugar_transpt.
DR   InterPro; IPR005829; Sug_transporter.
DR   Pfam; PF00083; sugar_tr; 1.
DR   PRINTS; PR00171; SUGRTRNSPORT.
DR   TIGRFAMs; TIGR00879; SP; 1.
DR   PROSITE; PS50850; MFS; 1.
DR   PROSITE; PS00216; SUGAR_TRANSPORT_1; 2.
DR   PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   498 AA;  54785 MW;  4458BF19DB84255D CRC64;

  Query Match             57.4%;  Score 1411.5;  DB 13;  Length 498;
  Best Local Similarity   58.0%;  Pred. No. 4.9e-87;
```

```
  Matches  282;  Conservative   63;  Mismatches  122;  Indels   19;  Gaps    5;

Qy          2 TPEDPEETQPLLGPPGGSAPRGRRVFLAAFAAALGPLSFGFALGYSSPAIPSLQRAAPPA 61
              | ||  |    |        :  ::|:| ||| |||||||| ||||||||| |:|
Db         17 TEEDRSEQDAYL-----DKVKNGKLFIATFAAVLGPLSFGFVLGYSSPAIPELRRIQDLR 71

Qy         62 PRLDDAAASWFGAVVTLGAAAGGVLGGWLVDRAGRKLSLLLCSVPFVAGFAVITAAQDVW 121
               :|    |||||:|||:||| ||:||||:|:| ||||||: |::||: ||  | |||: |
Db         72 LQLSVEEASWFGSVVTIGAALGGLLGGWIVERIGRKLSLMFCAIPFIFGFTTIIAAQNHW 131

Qy        122 MLLGGRLLTGLACGVASLVAPVYISEIAYPAVRGLLGSCVQLMVVVGILLAYLAGWVLEW 181
              |   ||:||||| || ||| |:||||:|:  ||| ||||||||||:||: ||: |  |:|
Db        132 MFYVGRVLTGLASGVTSLVVPLYISEMAHERVRGTLGSCVQLMVVIGIMGAYVTGLFLDW 191

Qy        182 RWLAVLGCVPPSLMLLLMCFMPETPRFLLTQHRRQEAMAALRFLWGSE--QGW-----ED 234
              |||||   :||:|||| |||||||||||| | :|:||  ||||| | :    |     ||
Db        192 RWLAVASSIPPTLMLLSMCFMPETPRFLLCQGKRREAEDALRFLRGPDAPAEWECARIED 251

Qy        235 PPIGAEQSFHLALLRQPGIYKPFIIGVSLMAFQQLSGVNAVMFYAETIFEEAKFKDSSLA 294
                   |||| |  |: ||:|||  ||| :|  || :|:||:|||||||||:| || | :|
Db        252 AYKNEEQSFSLGDLKDPGVYKPLGIGVMMMLLQQFTGINAIMFYAETIFEQAHFKSSDVA 311

Qy        295 SVVVGVIQVLFTAVAALIMDRAGRRLLLVLSGVVMVFSTSAFGAYFKLTQGGPGNSSHVA 354
              :|:|   ||:|||:||||||:|||::||:||||||  | : || |||||   | |||  :
Db        312 TVIVAATQVVFTAIAALIMDKAGRKVLLILSGVVMCVSEAVFGVYFKLTVMKPNNSSMTS 371

Qy        355 ISAP----VSAQPVDASVGLAWLAVGSMCLFIAGFAVGWGPIPWLLMSEIFPLHVKGVAT 410
              :       :  ||   |  |||||||||  ||||||:|||| |||:||||||  |:|: :
Db        372 VLTDTHGLLEDQP---SADLAWLAVGSMGFFIAGFAIGWGPTPWLVMSEIFPTRVRGLGS 428

Qy        411 GICVLTNWLMAFLVTKEFSSLMEVLRPYGAFWLASAFCIFSVLFTLFCVPETKGKTLEQI 470
               :||||||  ||:||| | :||: |   | ||: || |  :|:|| | ||||||||||:|
Db        429 ALCVLTNWTCAFIVTKTFQNLMDALSSAGTFWMFSALCASNVVFTAFFVPETKGKTLEEI 488

Qy        471 TAHFEG 476
               | |:|
Db        489 QAGFKG 494

RESULT 3
Q8AYP6
ID   Q8AYP6      PRELIMINARY;      PRT;   482 AA.
AC   Q8AYP6;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Glucose transporter type 8.
GN   GLUT8.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Seki Y., Sato K., Abe H., Akiba Y.;
RT   "Glocose transporter 8 plays a major role in insulin-responsive
RT   glucose transporter in place of glucose transporter 4 in chickens.";
RL   Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB083371; BAC20934.1; -.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0005351; F:sugar porter activity; IEA.
DR   GO; GO:0005215; F:transporter activity; IEA.
DR   GO; GO:0008643; P:carbohydrate transport; IEA.
DR   InterPro; IPR007114; MFS.
DR   InterPro; IPR005828; Sub_transporter.
DR   InterPro; IPR003663; Sugar_transpt.
DR   InterPro; IPR005829; Sug_transporter.
DR   Pfam; PF00083; sugar_tr; 1.
DR   PRINTS; PR00171; SUGRTRNSPORT.
DR   TIGRFAMs; TIGR00879; SP; 1.
DR   PROSITE; PS50850; MFS; 1.
DR   PROSITE; PS00216; SUGAR_TRANSPORT_1; 1.
```

```
DR   PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.
SQ   SEQUENCE   482 AA;  53349 MW;  CCFD0CA15F44C976 CRC64;

  Query Match             57.3%;  Score 1407;  DB 13;  Length 482;
  Best Local Similarity   59.2%;  Pred. No. 9.5e-87;
  Matches  282;  Conservative   68;  Mismatches  118;  Indels    8;  Gaps    3;

Qy          7 EETQPLLGPPGGSA---PRGRRVFLAAFAAALGPLSFGFALGYSSPAIPSLQRAAPPAPR 63
              ||:| ||      |    : ::::|| ||| |||||||| ||||||||| |::   |  |
Db          4 EESQYLLAHRTSDAYLRVQNKKLYLATFAAVLGPLSFGFVLGYSSPAIPELRKIDNPKLR 63

Qy         64 LDDAAASWFGAVVTLGAAAGGVLGGWLVDRAGRKLSLLLCSVPFVAGFAVITAAQDVWML 123
              ||   |||||::|||||||||:|||:|||: ||||||:|||:|||:|: || :||:||||
Db         64 LDSNQASWFGSIVTLGAAAGGILGGYLVDKIGRKLSLMLCSIPFVSGYIVIISAQNVWML 123

Qy        124 LGGRLLTGLACGVASLVAPVYISEIAYPAVRGLLGSCVQLMVVVGILLAYLAGWVLEWRW 183
                ||:||||| |: ||| |||||||::  |||:||| |||||| ||| ||:||  |:| |
Db        124 YFGRILTGLASGITSLVVPVYISEISHTDVRGMLGSFVQLMVVTGILGAYIAGLTLKWHW 183

Qy        184 LAVLGCVPPSLMLLLMCFMPETPRFLLTQHRRQEAMAALRFLWGS--EQGWEDPPIGA-- 239
              ||||   || :||| | |||||||||| | :| ||:|||:|| |   :  ||   | |
Db        184 LAVLCSFPPCVMLLFMLFMPETPRFLLDQKKRAEAIAALQFLRGPYVDHEWECRQIEANV 243

Qy        240 -EQSFHLALLRQPGIYKPFIIGVSLMAFQQLSGVNAVMFYAETIFEEAKFKDSSLASVVV 298
               |:   |   : | ||:| :||| ||  ||::|:|||| |||||||:| |:|| :|||||
Db        244 EEEGLSLFEFKNPSIYRPLLIGVILMFLQQVTGINAVMSYAETIFEDANFQDSRMASVVV 303

Qy        299 GVIQVLFTAVAALIMDRAGRRLLLVLSGVVMVFSTSAFGAYFKLTQGGPGNSSHVAISAP 358
              | ||| ||||||||:|: ||::|| :||::|  ||: || |||:      |||:  :
Db        304 GFIQVCFTAVAALIIDKTGRKVLLYVSGMIMALSTALFGFYFKMVLPNGNNSSNTDLWFT 363

Qy        359 VSAQPVDASVGLAWLAVGSMCLFIAGFAVGWGPIPWLLMSEIFPLHVKGVATGICVLTNW 418
              :::        |:|||| |: ||:||||:||||:|||:|||||||  :||::  ||||||
Db        364 LNSVTPGTDTRLSWLAVVSLGLFVAGFALGWGPVPWLVMSEIFPLKARGVSSSACVLTNW 423

Qy        419 LMAFLVTKEFSSLMEVLRPYGAFWLASAFCIFSVLFTLFCVPETKGKTLEQITAHF 474
              :|||||||||   :  |  || ||| ||||  || |  | ||||||:||||| |:|
Db        424 VMAFLVTKEFHDFIGFLTSYGTFWLFSAFCCLSVTFAAFYVPETKGRTLEQIEAYF 479


RESULT 4
Q8VDJ4
ID   Q8VDJ4      PRELIMINARY;      PRT;   266 AA.
AC   Q8VDJ4;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein (Fragment).
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Strausberg R.;
RL   Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
DR   EMBL; BC021758; AAH21758.1; -.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0005351; F:sugar porter activity; IEA.
DR   GO; GO:0005215; F:transporter activity; IEA.
DR   GO; GO:0008643; P:carbohydrate transport; IEA.
DR   InterPro; IPR007114; MFS.
DR   InterPro; IPR005828; Sub_transporter.
DR   InterPro; IPR003663; Sugar_transpt.
DR   InterPro; IPR005829; Sug_transporter.
DR   Pfam; PF00083; sugar_tr; 1.
DR   PRINTS; PR00171; SUGRTRNSPORT.
DR   PROSITE; PS50850; MFS; 1.
DR   PROSITE; PS00216; SUGAR_TRANSPORT_1; 1.
KW   Hypothetical protein; Sugar transport; Transmembrane; Transport.
FT   NON_TER       1      1
```

```
SQ   SEQUENCE   266 AA;  29132 MW;  1B3031928863B0FC CRC64;

  Query Match             47.5%;  Score 1167;  DB 11;  Length 266;
  Best Local Similarity   83.1%;  Pred. No. 6.9e-71;
  Matches  222;  Conservative   20;  Mismatches   23;  Indels    2;  Gaps    2;

Qy        212 QHRRQEAMAALRFLWGSEQGWEDPPIGAE-QSFHLALLRQPGIYKPFIIGVSLMAFQQLS 270
              ||: ||||||||||||||:|||:||:||| | | |||||:|||||| |||:||| |||||
Db          1 QHQYQEAMAALRFLWGSEEGWEEPPVGAEHQGFQLALLRRPGIYKPLIIGISLMVFQQLS 60

Qy        271 GVNAVMFYAETIFEEAKFKDSSLASVVVGVIQVLFTAVAALIMDRAGRRLLLVLSGVVMV 330
              ||||:|||| :||||||||||||||| ||:|||||||||||||||||||||| ||||:||
Db         61 GVNAIMFYANSIFEEAKFKDSSLASVTVGIIQVLFTAVAALIMDRAGRRLLLALSGVIMV 120

Qy        331 FSTSAFGAYFKLTQGGPGNSSHVAISAPVSAQPVDASVGLAWLAVGSMCLFIAGFAVGWG 390
              || |||| ||||||  | ||||| :  |::|:|||  |||||||||||||||||||||||
Db        121 FSMSAFGTYFKLTQSLPSNSSHVGL-VPIAAEPVDVQVGLAWLAVGSMCLFIAGFAVGWG 179

Qy        391 PIPWLLMSEIFPLHVKGVATGICVLTNWLMAFLVTKEFSSLMEVLRPYGAFWLASAFCIF 450
              |||||||||||||||||||||:|||||| |||||||||||:||:||||||||| :|||
Db        180 PIPWLLMSEIFPLHVKGVATGVCVLTNWFMAFLVTKEFSSVMEMLRPYGAFWLTAAFCAL 239

Qy        451 SVLFTLFCVPETKGKTLEQITAHFEGR 477
              ||||||  ||||||:||||:|||||||
Db        240 SVLFTLTVVPETKGRTLEQVTAHFEGR 266


RESULT 5
Q8WZ05
ID   Q8WZ05      PRELIMINARY;      PRT;   248 AA.
AC   Q8WZ05;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Zhang P.P., Zhou X.M., Jiang H.Q., Huang Y., Qin W.X., Zhao X.T.,
RA   Wan D.F., Gu J.R.;
RT   "Novel human cDNA clones with function of inhibiting cancer cell
RT   growth.";
RL   Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
DR   EMBL; AF289587; AAL55771.1; -.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0005215; F:transporter activity; IEA.
DR   GO; GO:0006810; P:transport; IEA.
DR   InterPro; IPR007114; MFS.
DR   InterPro; IPR005828; Sub_transporter.
DR   InterPro; IPR005829; Sug_transporter.
DR   Pfam; PF00083; sugar_tr; 1.
DR   PROSITE; PS50850; MFS; 1.
DR   PROSITE; PS00216; SUGAR_TRANSPORT_1; 1.
KW   Hypothetical protein; Transmembrane.
SQ   SEQUENCE   248 AA;  27017 MW;  AA929A749BBEE558 CRC64;

  Query Match             46.0%;  Score 1131;  DB 4;  Length 248;
  Best Local Similarity   93.0%;  Pred. No. 1.7e-68;
  Matches  227;  Conservative    5;  Mismatches    6;  Indels    6;  Gaps    2;

Qy        164 MVVVGILLAYLAGWVLEWRWLAVLGCVPPSLMLLLMCFMPETPRFLLTQHRRQEAMAALR 223
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MVVVGILLAYLAGWVLEWRWLAVLGCVPPSLMLLLMCFMPETPRFLLTQHRRQEAMAALR 60

Qy        224 FLWGSEQGWEDPPIGAEQSFHLALLRQPGIYKPFIIGVSLMAFQQLSGVNAVMFYAETIF 283
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 FLWGSEQGWEDPPIGAEQSFHLALLRQPGIYKPFIIGVSLMAFQQLSGVNAVMFYAETIF 120

Qy        284 EEAKFKDSSLASVVVGVIQVLFTAVAALIMDRAGRRLLLVLSGVVMVFSTSAFGAYFKLT 343
```

```
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 EEAKFKDSSLASVVVGVIQVLFTAVAALIMDRAGRRLLLVLSGVVMVFSTSAFGAYFKLT 180

Qy        344 QGGPGNSSHVAISAPVSAQPVDASVGLAWLAVGSMCLFIAGFAVGWGP-IPWLLMSEIFP 402
              |||||||||||||||||||||||||||||||||||||||||     ||   | |:: :
Db        181 QGGPGNSSHVAISAPVSAQPVDASVGLAWLAVGSMCLFIAG-----GPQALWSLLACLRF 235

Qy        403 LHVK 406
              ||::
Db        236 LHLQ 239

RESULT 6
Q8SPS1
ID   Q8SPS1      PRELIMINARY;      PRT;   246 AA.
AC   Q8SPS1;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Glucose transporter 8 (Fragment).
OS   Ovis aries (Sheep).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Caprinae; Ovis.
OX   NCBI_TaxID=9940;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Placenta;
RA   Limesand S.W., Regnault T.R.H., Hay W.W. Jr.;
RT   "Characterization of Glucose Transporter 8 (GLUT8) in the Ovine
RT   Placenta.";
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
DR   EMBL; AF495799; AAM18513.1; -.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0005351; F:sugar porter activity; IEA.
DR   GO; GO:0005215; F:transporter activity; IEA.
DR   GO; GO:0008643; P:carbohydrate transport; IEA.
DR   InterPro; IPR007114; MFS.
DR   InterPro; IPR005828; Sub_transporter.
DR   InterPro; IPR003663; Sugar_transpt.
DR   InterPro; IPR005829; Sug_transporter.
DR   Pfam; PF00083; sugar_tr; 1.
DR   PRINTS; PR00171; SUGRTRNSPORT.
DR   PROSITE; PS50850; MFS; 1.
DR   PROSITE; PS00216; SUGAR_TRANSPORT_1; 1.
KW   Transmembrane.
FT   NON_TER     246    246
SQ   SEQUENCE   246 AA;  26537 MW;  E8F3871946CD0723 CRC64;

  Query Match             45.8%;  Score 1125.5;  DB 6;  Length 246;
  Best Local Similarity   88.2%;  Pred. No. 3.9e-68;
  Matches  217;  Conservative   11;  Mismatches   17;  Indels    1;  Gaps    1;

Qy        219 MAALRFLWGSEQGWEDPPIGAE-QSFHLALLRQPGIYKPFIIGVSLMAFQQLSGVNAVMF 277
              |||::|||||  |||:|| ||| | ||:| || ||||||||||:||||||||||||||||
Db          1 MAAMQFLWGSAPGWEEPPTGAEHQGFHVAQLRCPGIYKPFIIGISLMAFQQLSGVNAVMF 60

Qy        278 YAETIFEEAKFKDSSLASVVVGVIQVLFTAVAALIMDRAGRRLLLVLSGVVMVFSTSAFG 337
              |||||||||||||||||||||||||||||| |||||||||||||| ||||||||||||||
Db         61 YAETIFEEAKFKDSSLASVVVGVIQVLFTATAALIMDRAGRRLLLTLSGVVMVFSTSAFG 120

Qy        338 AYFKLTQGGPGNSSHVAISAPVSAQPVDASVGLAWLAVGSMCLFIAGFAVGWGPIPWLLM 397
              ||||||:||| ||||| : |||| :| : :||||||||||||||||||||||||||||||
Db        121 AYFKLTEGGPSNSSHVDLPAPVSMEPANTNVGLAWLAVGSMCLFIAGFAVGWGPIPWLLM 180

Qy        398 SEIFPLHVKGVATGICVLTNWLMAFLVTKEFSSLMEVLRPYGAFWLASAFCIFSVLFTLF 457
              ||||||||||||||:|||||| ||||||||||||||||||||||||||||||| ||||| 
Db        181 SEIFPLHVKGVATGVCVLTNWFMAFLVTKEFSSLMEVLRPYGAFWLASAFCIFGVLFTLA 240

Qy        458 CVPETK 463
              ||||||
Db        241 CVPETK 246
```

```
RESULT 7
Q8BTN2
ID   Q8BTN2      PRELIMINARY;      PRT;   497 AA.
AC   Q8BTN2;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Similar to solute carrier family 2.
GN   F630103L12RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NOD;
RX   MEDLINE=22354683; PubMed=12466851;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
DR   EMBL; AK089246; BAC40811.1; -.
DR   MGD; MGI:2443286; F630103L12Rik.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0005351; F:sugar porter activity; IEA.
DR   GO; GO:0005215; F:transporter activity; IEA.
DR   GO; GO:0008643; P:carbohydrate transport; IEA.
DR   InterPro; IPR007114; MFS.
DR   InterPro; IPR005828; Sub_transporter.
DR   InterPro; IPR003663; Sugar_transpt.
DR   InterPro; IPR005829; Sug_transporter.
DR   Pfam; PF00083; sugar_tr; 1.
DR   PRINTS; PR00171; SUGRTRNSPORT.
DR   TIGRFAMs; TIGR00879; SP; 1.
DR   PROSITE; PS50850; MFS; 1.
DR   PROSITE; PS00216; SUGAR_TRANSPORT_1; 1.
DR   PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.
SQ   SEQUENCE   497 AA;  54445 MW;  8E14D1DCD4B244E3 CRC64;

  Query Match             38.5%;  Score 946;  DB 11;  Length 497;
  Best Local Similarity   43.6%;  Pred. No. 9.6e-56;
  Matches  214;  Conservative   77;  Mismatches  168;  Indels   32;  Gaps    9;

Qy          3 PEDPEETQPLLGPPG----GSAPRGRRVFLAAFAAALGPLSFGFALGYSSPAIPSLQRAA 58
              || |         ||      | : |||||| ||| ||  |||:|| |:|| || |: ::
Db         17 PEVP-------ATPGERERAGALKNRRVFLATFAAVLGNFSFGYALVYTSPVIPELKLSS 69

Qy         59 PPAPRLDDAAASWFGAVVTLGAAAGGVLGGWLVDRAGRKLSLLLCSVPFVAGFAVITAAQ 118
               ||  ||   |||||:| ||||||||:    | |  |||||::  :||   |:|::  |:
Db         70 DPALHLDKIQASWFGSVFTLGAAAGGLSAMLLNDLLGRKLSIMFSAVPSAIGYAIMAGAR 129

Qy        119 DVWMLLGGRLLTGLACGVASLVAPVYISEIAYPAVRGLLGSCVQLMVVVGILLAYLAGWV 178
               :|||| ||:||| | |: :   |||:|||| | ||| ||:  ||| | | |  |  | :
Db        130 GLWMLLLGRMLTGFAGGLTAACIPVYVSEIAPPDVRGALGATPQLMAVFGSLSLYALGLL 189

Qy        179 LEWRWLAVLGCVPPSLMLLLMCFMPETPRFLLTQHRRQEAMAALRFLWG-SEQGWEDPPI 237
              | |||||| |  |  :|:||: ||| :|||||:: | :||: || :|   ||  ||   |
Db        190 LPWRWLAVAGEGPVLIMILLLSFMPNSPRFLLSKSRDEEALQALTWLRADSEVHWEFEQI 249

Qy        238 G---AEQSFHL--ALLRQPGIYKPFIIGVSLMAFQQLSGVNAVMFYAETIFEEAK-FKDS 291
                    ||  :  |  |:| :|:| :| | :   |||:|:  :: | :|||:       |
Db        250 QDNVRRQSSRVSWAEAREPRVYRPVLIAVLMRFLQQLTGITPILVYLQTIFDNTSVVLPS 309

Qy        292 SLASVVVGVIQVLFTAVAALIMDRAGRRLLLVLSGVVMVFSTSAFGAYFK-----LTQGG 346
                 : :|| :::|   :||: || |||::|| :|  ||  :    | | :     ||
Db        310 QQDAAIVGAVRLLSVLIAAVTMDLAGRKVLLYVSASVMFAANLTLGLYVQFVPRPLT--- 366

Qy        347 PGNSSHVAISAPVSAQPVDASVGLAWLAVGSMCLFIAGFAVGWGPIPWLLMSEIFPLHVK 406
              | ::  : |          |   |  : : :  ||| |:|:||||| ||||||: ||  :
Db        367 PNSTVEIVILGDT------AFNYLTLIPLLATMLFIMGYAMGWGPITWLLMSEVLPLRAR 420
```

```
Qy    407 GVATGICVLTNWLMAFLVTKEFSSLMEVLRPYGAFWLASAFCIFSVLFTLFCVPETKGKT 466
          |||:|:||| :|| ||::|  |    :       |:  || |: |:|||  |||||:|::
Db    421 GVASGLCVLVSWLTAFVLTNYFLLAVNAFGLQVPFFFFSAICLLSLLFTGCCVPETRGRS 480

Qy    467 LEQITAHFEGR 477
          |||| | |  |
Db    481 LEQIEAFFHTR 491

RESULT 8
Q8NCC2
ID   Q8NCC2      PRELIMINARY;      PRT;   445 AA.
AC   Q8NCC2;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein FLJ90355.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Isogai T., Ota T., Nishikawa T., Hayashi K., Otsuki T., Sugiyama T.,
RA   Suzuki Y., Nagai K., Sugano S., Ishii S., Kawai-Hio Y., Saito K.,
RA   Yamamoto J., Wakamatsu A., Nakamura Y., Kojima S., Nagahari K.,
RA   Masuho Y., Ono T., Okano K., Yoshikawa Y., Aotsuka S., Sasaki N.,
RA   Hattori A., Okumura K., Iwayanagi T., Ninomiya K.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
DR   EMBL; AK074836; BAC11235.1; -.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0005351; F:sugar porter activity; IEA.
DR   GO; GO:0005215; F:transporter activity; IEA.
DR   GO; GO:0008643; P:carbohydrate transport; IEA.
DR   InterPro; IPR007114; MFS.
DR   InterPro; IPR005828; Sub_transporter.
DR   InterPro; IPR003663; Sugar_transpt.
DR   InterPro; IPR005829; Sug_transporter.
DR   Pfam; PF00083; sugar_tr; 1.
DR   PRINTS; PR00171; SUGRTRNSPORT.
DR   PROSITE; PS50850; MFS; 1.
DR   PROSITE; PS00216; SUGAR_TRANSPORT_1; 1.
DR   PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.
KW   Hypothetical protein; Sugar transport; Transmembrane; Transport.
SQ   SEQUENCE   445 AA;  48040 MW;  BF37DC0C313A32CE CRC64;

  Query Match             36.4%;  Score 894;  DB 4;  Length 445;
  Best Local Similarity   40.1%;  Pred. No. 2.7e-52;
  Matches  200;  Conservative   73;  Mismatches  134;  Indels   92;  Gaps    9;

Qy     10 QPLLG------------PPGGSAPRGR-------RVFLAAFAAALGPLSFGFALGYSSPA 50
          :||||            ||     | |       ||||| ||| ||  |||:|| |:||
Db      3 EPLLGAEGPDYDTFPEKPPPSPGDRARVGTLQNKRVFLATFAAVLGNFSFGYALVYTSPV 62

Qy     51 IPSLQRAAPPAPRLDDAAASWFGAVVTLGAAAGGVLGGWLVDRAGRKLSLLLCSVPFVAG 110
          ||:|:|:  |   |  : |||||:| ||||||||:    | |  |||||::  :||  ||
Db     63 IPALERSLDPDLHLTKSQASWFGSVFTLGAAAGGLSAMILNDLLGRKLSIMFSAVPSAAG 122

Qy    111 FAVITAAQDVWMLLGGRLLTGLACGVASLVAPVYISEIAYPAVRGLLGSCVQLMVVVGIL 170
          :|::  |  :|||| || ||| | |: :   |||:|||| | ||| ||:  ||| | | |
Db    123 YALMAGAHGLWMLLLGRTLTGFAGGLTAACIPVYVSEIAPPGVRGALGATPQLMAVFGSL 182

Qy    171 LAYLAGWVLEWRWLAVLGCVPPSLMLLLMCFMPETPRFLLTQHRRQEAMAALRFLWGSEQ 230
            |  | :| |||||| |  |  :|:||: ||| :|||||:: | :||: || :| |::
Db    183 SLYALGLLLPWRWLAVAGEAPVLIMILLLSFMPNSPRFLLSRGRDEEALRALAWLRGTDV 242

Qy    231 G--WEDPPIG---AEQSFHL--ALLRQPGIYKPFIIGVSLMAFQQLSGVNAVMFYAETIF 283
             ||   |      ||  :  |  | | : :|  : : :   |||:|:  :: | ::||
Db    243 DVHWEFEQIQDNVRRQSSRVSWAEARAPHVCRPITVALLMRLLQQLTGITPILVYLQSIF 302
```

```
Qy    284 EEAKF----KDSSLASVVVGVIQVLFTAVAALIMDRAGRRLLLVLSGVVMVFSTSAFGAY 339
          :        ||    : :|| :::|   :||| || |||::|
Db    303 DSTAVLLPPKDD---AAIVGAVRLLSVLIAALTMDLAGRKVL------------------ 341

Qy    340 FKLTQGGPGNSSHVAISAPVSAQPVDASVGLAWLAVGSMCLFIAGFAVGWGPIPWLLMSE 399
                                                  ||::|:||||||| ||||||
Db    342 ---------------------------------------LFVSGYAVGWGPITWLLMSE 361

Qy    400 IFPLHVKGVATGICVLTNWLMAFLVTKEFSSLMEVLRPYGAFWLASAFCIFSVLFTLFCV 459
          : ||  :|||:|:||| :|| ||::|| |  ::        |:  :| |: |::||  ||
Db    362 VLPLRARGVASGLCVLASWLTAFVLTKSFLPVVSTFGLQVPFFFFAAICLVSLVFTGCCV 421

Qy    460 PETKGKTLEQITAHFE-GR 477
          |||||::|||| : |  ||
Db    422 PETKGRSLEQIESFFRTGR 440

RESULT 9
Q8MKK4
ID   Q8MKK4      PRELIMINARY;      PRT;   488 AA.
AC   Q8MKK4;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   AT19440p (CG8234-PA).
GN   CG8234.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA   Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA   George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA   Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
RA   Patel S., Phouanenavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA   Celniker S.;
RL   Submitted (JUN-2002) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
```

```
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA   Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA   Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA   Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA   Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA   Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA   Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA   Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA   McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA   Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA   Phouanenavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA   Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA   Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT   "Sequencing of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA   Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA   Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA   Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA   Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA   Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA   Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT   "Annotation of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RA   Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   SEQUENCE FROM N.A.
RA   FlyBase;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
DR   EMBL; AY122079; AAM52591.1; -.
DR   EMBL; AE003825; AAM68715.1; -.
DR   FlyBase; FBgn0033644; CG8234.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0005351; F:sugar porter activity; IEA.
DR   GO; GO:0005215; F:transporter activity; IEA.
DR   GO; GO:0008643; P:carbohydrate transport; IEA.
DR   InterPro; IPR007114; MFS.
DR   InterPro; IPR005828; Sub_transporter.
DR   InterPro; IPR003663; Sugar_transpt.
DR   InterPro; IPR005829; Sug_transporter.
DR   Pfam; PF00083; sugar_tr; 1.
DR   PRINTS; PR00171; SUGRTRNSPORT.
DR   TIGRFAMs; TIGR00879; SP; 1.
DR   PROSITE; PS50850; MFS; 1.
DR   PROSITE; PS00216; SUGAR_TRANSPORT_1; 2.
DR   PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.
KW   Sugar transport; Transmembrane; Transport.
SQ   SEQUENCE   488 AA;  53088 MW;  6554C96EBEA1C08B CRC64;

  Query Match             32.7%;  Score 804.5;  DB 5;  Length 488;
  Best Local Similarity   37.4%;  Pred. No. 3.1e-46;
  Matches  173;  Conservative   81;  Mismatches  177;  Indels   31;  Gaps    5;

Qy         28 LAAFAAALGPLSFGFALGYSSPAIPSLQRAAPPAPRLDDAAASWFGAVVTLGAAAGGVLG 87
              ||| : :|  |  ||   |:|||: |:      :  :   | || | :: | | |||: |
Db         30 LAALSVSLCSLVVGFVSAYTSPALVSMTDRTITSFEVTKDAGSWVGGIMPLAALAGGITG 89

Qy         88 GWLVDRAGRKLSLLLCSVPFVAGFAVITAAQDVWMLLGGRLLTGLACGVASLVAPVYISE 147
              | |::  ||: ::|  :|||:    :|  | :| |:| || |||   |:|||  |||: |
Db         90 GPLIEYLGRRSTILATAVPFIVSSLLIACAVNVIMILCGRFLTGFCVGIASLSLPVYLGE 149

Qy        148 IAYPAVRGLLGSCVQLMVVVGILLAYLAGWVLEWRWLAVLGCVPPSLMLLLMCFMPETPR 207
                 | ||| ||     :  :|||: |:||  : |  || ||   |   |:||  :|||||
Db        150 TLQPEVRGTLGLLPTALGNIGILVCYVAGSFMNWSMLAFLGAALPVPFLILMIIIPETPR 209

Qy        208 FLLTQHRRQEAMAALRFLWGSE-----------QGWEDPPIGAEQSFHLALLRQPGIYKP 256
              : : : : : |  ||::| | |           |   |    | |:  | | ::  : ||
Db        210 WFVNRGQEERARKALKWLRGKEADVEPELKELMQSQADADRQATQNTCLELFKRNNL-KP 268

Qy        257 FIIGVSLMAFQQLSGVNAVMFYAETIFEEA-KFKDSSLASVVVGVIQVLFTAVAALIMDR 315
                | : || ||| ||:|||:||   ||::|    ||:|::::|||:    | :  :::||
Db        269 LSISLGLMFFQQFSGINAVIFYTVQIFKDAGSTIDSNLSTIIVGVVNFFATFMGIILIDR 328

Qy        316 AGRRLLLVLSGVVMVFSTSAFGAYFKLTQGGPGNSSHVAISAPVSAQPVDASVGLAWLAV 375
               ||::|| :| : |: : |  |: |     || : ||                 | || :
Db        329 LGRKILLYVSDIAMIVTLSILGGFFYCKAHGP-DVSH-----------------LGWLPL 370

Qy        376 GSMCLFIAGFAVGWGPIPWLLMSEIFPLHVKGVATGICVLTNWLMAFLVTKEFSSLMEVL 435
                  ::| ||::|:||||||:| || |  ::| | :     ||   |:||| |  |   :
Db        371 TCFVIYILGFSLGFGPIPWLMMGEILPAKIRGPAASVVTAFNWFCTFVVTKTFQDLTVAM 430

Qy        436 RPYGAFWLASAFCIFSVLFTLFCVPETKGKTLEQITAHFEGR 477
                :|||||  | ||  : | :  ||||:||:||:|     ||
Db        431 GAHGAFWLFGAICIVGLFFVIIFVPETRGKSLEEIERKMMGR 472

RESULT 10
Q86P59
ID   Q86P59      PRELIMINARY;      PRT;   506 AA.
AC   Q86P59;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   RH04286p.
GN   CG30035.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=y;
RA   Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA   Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA   George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA   Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
RA   Patel S., Phouanenavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA   Celniker S.;
RL   Submitted (FEB-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BT003466; AAO39469.1; -.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0005351; F:sugar porter activity; IEA.
DR   GO; GO:0005215; F:transporter activity; IEA.
DR   GO; GO:0008643; P:carbohydrate transport; IEA.
DR   InterPro; IPR007114; MFS.
DR   InterPro; IPR005828; Sub_transporter.
DR   InterPro; IPR003663; Sugar_transpt.
DR   InterPro; IPR005829; Sug_transporter.
DR   Pfam; PF00083; sugar_tr; 1.
DR   PRINTS; PR00171; SUGRTRNSPORT.
DR   TIGRFAMs; TIGR00879; SP; 1.
DR   PROSITE; PS50850; MFS; 1.
DR   PROSITE; PS00216; SUGAR_TRANSPORT_1; 2.
DR   PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.
SQ   SEQUENCE   506 AA;  55235 MW;  24E8718260D4319B CRC64;
```

```
  Query Match             32.1%;  Score 789.5;  DB 5;  Length 506;
  Best Local Similarity   37.4%;  Pred. No. 3.3e-45;
  Matches  173;  Conservative   75;  Mismatches  183;  Indels   31;  Gaps    5;

Qy           28 LAAFAAALGPLSFGFALGYSSPAIPSLQRAAPPAPRLDDAAASWFGAVVTLGAAAGGVLG 87
                ||| : :|| |  ||   |:|||: |:      :  :   | || | :: |   |||: |
Db           47 LAALSVSLGSLVVGFVSAYTSPALVSMTDRNITSFEVTQDAGSWVGGIMPLAGLAGGIAG 106

Qy           88 GWLVDRAGRKLSLLLCSVPFVAGFAVITAAQDVWMLLGGRLLTGLACGVASLVAPVYISE 147
                | |::  ||: ::|  :|||:    :|  | :| |:| || | |   |:|||  |||: |
Db          107 GPLIEYLGRRNTILATAVPFIVSSLLIACAVNVAMVLCGRFLAGFCVGIASLSLPVYLGE 166

Qy          148 IAYPAVRGLLGSCVQLMVVVGILLAYLAGWVLEWRWLAVLGCVPPSLMLLLMCFMPETPR 207
                   | ||| ||        :|||| ::||  : |  || ||   |   |:||  :|||||
Db          167 TVQPEVRGTLGLLPTAFGNIGILLCFVAGSFMNWSMLAFLGAALPVPFLILMFLIPETPR 226

Qy          208 FLLTQHRRQEAMAALRFLWGSEQGWE-----------DPPIGAEQSFHLALLRQPGIYKP 256
                : : :   : |  ||::| | |   |           |    | ::  | ||:   : ||
Db          227 WFVGRGLEERARKALKWLRGKEADVEPELKGLMRSQADADRQASRNTMLELLKLNNL-KP 285

Qy          257 FIIGVSLMAFQQLSGVNAVMFYAETIFEEA-KFKDSSLASVVVGVIQVLFTAVAALIMDR 315
                  | : || ||| ||:|||:||   ||::|    | :| :|:||::  | | :  :::||
Db          286 LSISLGLMFFQQFSGINAVIFYTVQIFKDAGSTIDGNLCTVIVGIVNFLATFIGIVLIDR 345

Qy          316 AGRRLLLVLSGVVMVFSTSAFGAYFKLTQGGPGNSSHVAISAPVSAQPVDASVGLAWLAV 375
                |||::|| :| : || :    | :|     || : ||                 | || :
Db          346 AGRKILLYVSDIAMVLTLFVLGGFFYCKTYGP-DVSH-----------------LGWLPL 387

Qy          376 GSMCLFIAGFAVGWGPIPWLLMSEIFPLHVKGVATGICVLTNWLMAFLVTKEFSSLMEVL 435
                    ::| ||::|:||||||:| || |  ::| |  :    ||   |:||| |  |   :
Db          388 TCFVIYILGFSLGFGPIPWLMMGEILPAKIRGSAASVATAFNWFCTFVVTKTFQDLTVAM 447

Qy          436 RPYGAFWLASAFCIFSVLFTLFCVPETKGKTLEQITAHFEGR 477
                  :|||||  | |   : | :  ||||:||||| |     ||
Db          448 GAHGAFWLFGAICFVGLFFVIIYVPETQGKTLEDIERKMMGR 489


RESULT 11
Q9V609
ID   Q9V609      PRELIMINARY;      PRT;   489 AA.
AC   Q9V609;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   CG7801 protein.
GN   CG30035 OR CG7797 OR CG7801.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
DR   EMBL; AE003825; AAF58631.1; -.
DR   FlyBase; FBgn0050035; CG30035.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0005351; F:sugar porter activity; IEA.
DR   GO; GO:0005215; F:transporter activity; IEA.
DR   GO; GO:0008643; P:carbohydrate transport; IEA.
DR   InterPro; IPR007114; MFS.
DR   InterPro; IPR005828; Sub_transporter.
DR   InterPro; IPR003663; Sugar_transpt.
DR   InterPro; IPR005829; Sug_transporter.
DR   Pfam; PF00083; sugar_tr; 1.
DR   PRINTS; PR00171; SUGRTRNSPORT.
DR   TIGRFAMs; TIGR00879; SP; 1.
DR   PROSITE; PS50850; MFS; 1.
DR   PROSITE; PS00216; SUGAR_TRANSPORT_1; 2.
DR   PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.
KW   Sugar transport; Transmembrane; Transport.
SQ   SEQUENCE   489 AA;  53026 MW;  D26436DAD69723C6 CRC64;

  Query Match             32.1%;  Score 788.5;  DB 5;  Length 489;
  Best Local Similarity   37.2%;  Pred. No. 3.7e-45;
  Matches  172;  Conservative   76;  Mismatches  183;  Indels   31;  Gaps    5;

Qy           28 LAAFAAALGPLSFGFALGYSSPAIPSLQRAAPPAPRLDDAAASWFGAVVTLGAAAGGVLG 87
                ||| : :|| |  ||   |:|||: |:      :  :   | || | :: |   |||: |
Db           30 LAALSVSLGSLVVGFVSAYTSPALVSMTDRNITSFEVTQDAGSWVGGIMPLAGLAGGIAG 89

Qy           88 GWLVDRAGRKLSLLLCSVPFVAGFAVITAAQDVWMLLGGRLLTGLACGVASLVAPVYISE 147
                | |::  ||: ::|  :|||:    :|  | :| |:| || | |   |:|||  |||: |
Db           90 GPLIEYLGRRNTILATAVPFIVSSLLIACAVNVAMVLCGRFLAGFCVGIASLSLPVYLGE 149

Qy          148 IAYPAVRGLLGSCVQLMVVVGILLAYLAGWVLEWRWLAVLGCVPPSLMLLLMCFMPETPR 207
                   | ||| ||        :|||| ::||  : |  || ||   |   |:||  :|||||
Db          150 TVQPEVRGTLGLLPTAFGNIGILLCFVAGSFMNWSMLAFLGAALPVPFLILMFLIPETPR 209

Qy          208 FLLTQHRRQEAMAALRFLWGSEQGWE-----------DPPIGAEQSFHLALLRQPGIYKP 256
                : : :   : |  ||::| | |   |           |    | ::  | ||:   : ||
Db          210 WFVGRGLEERARKALKWLRGKEADVEPELKGLMRSQADADRQASRNTMLELLKLNNL-KP 268

Qy          257 FIIGVSLMAFQQLSGVNAVMFYAETIFEEA-KFKDSSLASVVVGVIQVLFTAVAALIMDR 315
                  | : || ||| ||:|||:||   ||::|    | :| :::||::  | | :  :::||
Db          269 LSISLGLMFFQQFSGINAVIFYTVQIFKDAGSTIDGNLCTIIVGIVNFLATFIGIVLIDR 328

Qy          316 AGRRLLLVLSGVVMVFSTSAFGAYFKLTQGGPGNSSHVAISAPVSAQPVDASVGLAWLAV 375
                |||::|| :| : || :    | :|     || : ||                 | || :
Db          329 AGRKILLYVSDIAMVLTLFVLGGFFYCKTYGP-DVSH-----------------LGWLPL 370

Qy          376 GSMCLFIAGFAVGWGPIPWLLMSEIFPLHVKGVATGICVLTNWLMAFLVTKEFSSLMEVL 435
                    ::| ||::|:||||||:| || |  ::| |  :    ||   |:||| |  |   :
Db          371 TCFVIYILGFSLGFGPIPWLMMGEILPAKIRGSAASVATAFNWFCTFVVTKTFQDLTVAM 430
```

```
Qy        436 RPYGAFWLASAFCIFSVLFTLFCVPETKGKTLEQITAHFEGR 477
                :|||||  | |   : | :  ||||:||||| |     ||
Db        431 GAHGAFWLFGAICFVGLFFVIIYVPETQGKTLEDIERKMMGR 472

RESULT 12
Q9V608
ID   Q9V608      PRELIMINARY;      PRT;   857 AA.
AC   Q9V608;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   CG30035 protein.
GN   CG30035 OR CG7797 OR CG7801.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA   Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA   Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA   Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA   Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA   Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA   Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA   Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA   McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA   Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA   Phouanenavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA   Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA   Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT   "Sequencing of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA   Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA   Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA   Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA   Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA   Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA   Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT   "Annotation of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RA   FlyBase;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
DR   EMBL; AE003825; AAF58632.2; -.
DR   FlyBase; FBgn0050035; CG30035.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0005351; F:sugar porter activity; IEA.
DR   GO; GO:0005215; F:transporter activity; IEA.
DR   GO; GO:0008643; P:carbohydrate transport; IEA.
DR   InterPro; IPR007114; MFS.
DR   InterPro; IPR005828; Sub_transporter.
DR   InterPro; IPR003663; Sugar_transpt.
DR   InterPro; IPR005829; Sug_transporter.
DR   Pfam; PF00083; sugar_tr; 1.
DR   PRINTS; PR00171; SUGRTRNSPORT.
DR   TIGRFAMs; TIGR00879; SP; 1.
DR   PROSITE; PS50850; MFS; 1.
DR   PROSITE; PS00216; SUGAR_TRANSPORT_1; 2.
DR   PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.
KW   Sugar transport; Transmembrane; Transport.
SQ   SEQUENCE   857 AA;  95188 MW;  8408E1191B8B7A5F CRC64;

  Query Match             32.1%;  Score 788.5;  DB 5;  Length 857;
  Best Local Similarity   37.2%;  Pred. No. 6.6e-45;
  Matches  172;  Conservative   76;  Mismatches  183;  Indels   31;  Gaps    5;

Qy         28 LAAFAAALGPLSFGFALGYSSPAIPSLQRAAPPAPRLDDAAASWFGAVVTLGAAAGGVLG 87
              ||| : :|| |  ||   |:|||: |:      :  :   | || | :: |   |||: |
Db        398 LAALSVSLGSLVVGFVSAYTSPALVSMTDRNITSFEVTQDAGSWVGGIMPLAGLAGGIAG 457

Qy         88 GWLVDRAGRKLSLLLCSVPFVAGFAVITAAQDVWMLLGGRLLTGLACGVASLVAPVYISE 147
              | |::  ||: ::|  :|||:    :|  | :| |:| || | |   |:|||  |||: |
Db        458 GPLIEYLGRRNTILATAVPFIVSSLLIACAVNVAMVLCGRFLAGFCVGIASLSLPVYLGE 517

Qy        148 IAYPAVRGLLGSCVQLMVVVGILLAYLAGWVLEWRWLAVLGCVPPSLMLLLMCFMPETPR 207
                 | ||| ||        :|||| ::||  : |  || ||   |   |:|   :|||||
Db        518 TVQPEVRGTLGLLPTAFGNIGILLCFVAGSFMNWSMLAFLGAALPVPFLILMFLIPETPR 577

Qy        208 FLLTQHRRQEAMAALRFLWGSEQGWE-----------DPPIGAEQSFHLALLRQPGIYKP 256
              : : :   : |  ||::| | |   |           |    | ::  | ||:   : ||
Db        578 WFVGRGLEERARKALKWLRGKEADVEPELKGLMRSQADADRQASRNTMLELLKLNNL-KP 636

Qy        257 FIIGVSLMAFQQLSGVNAVMFYAETIFEEA-KFKDSSLASVVVGVIQVLFTAVAALIMDR 315
                | : || ||| ||:|||:||   ||::|    | :| :::||::  | | :  :::||
Db        637 LSISLGLMFFQQFSGINAVIFYTVQIFKDAGSTIDGNLCTIIVGIVNFLATFIGIVLIDR 696

Qy        316 AGRRLLLVLSGVVMVFSTSAFGAYFKLTQGGPGNSSHVAISAPVSAQPVDASVGLAWLAV 375
              |||::|| :| : || :    | :|     || : ||                 | || :
Db        697 AGRKILLYVSDIAMVLTLFVLGGFFYCKTYGP-DVSH-----------------LGWLPL 738
```

```
Qy          376 GSMCLFIAGFAVGWGPIPWLLMSEIFPLHVKGVATGICVLTNWLMAFLVTKEFSSLMEVL 435
                    ::| ||::|:||||||:| || |  ::| |  :    ||   |:||| |  |   :
Db          739 TCFVIYILGFSLGFGPIPWLMMGEILPAKIRGSAASVATAFNWFCTFVVTKTFQDLTVAM 798

Qy          436 RPYGAFWLASAFCIFSVLFTLFCVPETKGKTLEQITAHFEGR 477
                  :|||||  | |   : | :  ||||:||||| |     ||
Db          799 GAHGAFWLFGAICFVGLFFVIIYVPETQGKTLEDIERKMMGR 840

RESULT 13
Q9VU17
ID   Q9VU17      PRELIMINARY;      PRT;   539 AA.
AC   Q9VU17;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   CG10960 protein (HL01062p).
GN   CG10960.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RA   Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA   Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA   George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA   Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
RA   Patel S., Phouanenavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
```

```
RA   Celniker S.;
RL   Submitted (JUN-2002) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
DR   EMBL; AE003540; AAF49874.1; -.
DR   EMBL; AY119564; AAM50218.1; -.
DR   FlyBase; FBgn0036316; CG10960.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0005351; F:sugar porter activity; IEA.
DR   GO; GO:0005215; F:transporter activity; IEA.
DR   GO; GO:0008643; P:carbohydrate transport; IEA.
DR   InterPro; IPR007114; MFS.
DR   InterPro; IPR005828; Sub_transporter.
DR   InterPro; IPR003663; Sugar_transpt.
DR   InterPro; IPR005829; Sug_transporter.
DR   Pfam; PF00083; sugar_tr; 1.
DR   PRINTS; PR00171; SUGRTRNSPORT.
DR   TIGRFAMs; TIGR00879; SP; 1.
DR   PROSITE; PS50850; MFS; 1.
DR   PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.
KW   Sugar transport; Transmembrane; Transport.
SQ   SEQUENCE   539 AA;  58413 MW;  A2F4814DBF676BBD CRC64;

  Query Match             32.1%;  Score 788;  DB 5;  Length 539;
  Best Local Similarity   36.2%;  Pred. No. 4.4e-45;
  Matches  178;  Conservative   95;  Mismatches  183;  Indels   36;  Gaps    6;

Qy            2 TPEDPEETQPLLGPPGGSAPRGRR----VFLAAFAAALGPLSFGFALGYSSPAIPSL-QR 56
                |   | : : |||    ::          ::|| ||| |  : |  ||::|||   :  |
Db           55 TSSTPHDQENLLGNMVHTSTSNNSQKWPQYVAALAAAGGAFAAGTVLGWTSPAETEIVDR 114

Qy           57 AAPPAPRLDDAAASWFGAVVTLGAAAGGVLGGWLVDRAGRKLSLLLCSVPFVAGFAVITA 116
                       :|    || |: :|||||   :  |:|::  |||  ::|   :||: |: ::
Db          115 GEGYDFPVDKDQFSWVGSAMTLGAACVCIPIGFLINMIGRKWTMLFLVLPFILGWTMLIW 174

Qy          117 AQDVWMLLGGRLLTGLACGVASLVAPVYISEIAYPAVRGLLGSCVQLMVVVGILLAYLAG 176
                | :| ||   | : |:| |   : ||:|  |||   :|| |||  |||: :|||  |  |
Db          175 AVNVSMLYASRFILGIAGGAFCVTAPMYTGEIAQKEIRGTLGSFFQLMITIGILFVYAVG 234

Qy          177 WVLEWRWLAVLGCVPPSLMLLLMCFMPETPRFLLTQHRRQEAMAALRFLWGSEQGWE--- 233
                  ::  ||:::  : | :   :  ||||:| :|::: | : |: ::::| | |  :|
Db          235 AGVKIFWLSIICGILPLIFGAIFFFMPESPTYLVSKDRSENAIKSIQWLRGKEYDYEPEL 294

Qy          234 ------DPPIGAEQSFHLALLRQPGIYKPFIIGVSLMAFQQLSGVNAVMFYAETIFEEAK 287
                      |    | :    | | :|   |   | : ||  |||: |:|||:|||  || ||
Db          295 AELRETDRETKANKVNVWAALNRPVTRKALAISMGLMFFQQVCGINAVIFYASRIFLEAN 354

Qy          288 FK-DSSLASVVVGVIQVLFTAVAALIMDRAGRRLLLVLSGVVMVFSTSAFGAYFKLTQGG 346
                   ::  |::::|::||: | |: |::|: |||:||: ||: |  ||:| | || |
Db          355 TGIEAEWATILIGIMQVVATFVSTLVVDKLGRRILLLASGISMAISTTAIGVYFFL---- 410

Qy          347 PGNSSHVAISAPVSAQPVDAS--VGLAWLAVGSMCLFIAGFAVGWGPIPWLLMSEIFPLH 404
                               |  ||:  | | || | |:||||  |::|:||:|||:| |:|
Db          411 ---------------QKQDAAQVVSLGWLPVASLCLFIIMFSIGYGPVPWLMMGELFATD 455

Qy          405 VKGVATGICVLTNWLMAFLVTKEFSSLMEVLRPYGAFWLASAFCIFSVLFTLFCVPETKG 464
                :|| |  :   :|||:||:||| | :| : |   | ||| :   :  |:|  | ||||||
Db          456 IKGFAGSLAGTSNWLLAFVVTKTFVNLNDGLGIGGTFWLFAGLTVVGVIFVYFAVPETKG 515

Qy          465 KTLEQITAHFEG 476
                |:| :|     |
Db          516 KSLNEIQQELAG 527

RESULT 14
Q8IQH6
ID   Q8IQH6      PRELIMINARY;      PRT;   471 AA.
AC   Q8IQH6;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   CG10960-PA.
GN   CG10960.
```

```
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA   Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., WoodageT, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA   Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA   Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA   Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA   Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA   Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA   Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA   Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA   Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA   McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA   Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA   Phouanenavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA   Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA   Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT   "Sequencing of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA   Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA   Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA   Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA   Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA   Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA   Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT   "Annotation of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RA   FlyBase;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003540; AAN11863.1; -.
DR   FlyBase; FBgn0036316; CG10960.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0005351; F:sugar porter activity; IEA.
DR   GO; GO:0005215; F:transporter activity; IEA.
DR   GO; GO:0008643; P:carbohydrate transport; IEA.
DR   InterPro; IPR007114; MFS.
DR   InterPro; IPR005828; Sub_transporter.
DR   InterPro; IPR003663; Sugar_transpt.
DR   InterPro; IPR005829; Sug_transporter.
DR   Pfam; PF00083; sugar_tr; 1.
DR   PRINTS; PR00171; SUGRTRNSPORT.
DR   TIGRFAMs; TIGR00879; SP; 1.
DR   PROSITE; PS50850; MFS; 1.
DR   PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.
SQ   SEQUENCE   471 AA;  51323 MW;  33199A53BC0EF65B CRC64;

  Query Match             32.0%;  Score 787;  DB 5;  Length 471;
  Best Local Similarity   37.4%;  Pred. No. 4.4e-45;
  Matches  173;  Conservative   91;  Mismatches  167;  Indels   32;  Gaps    5;

Qy         27 FLAAFAAALGPLSFGFALGYSSPAIPSL-QRAAPPAPRLDDAAASWFGAVVTLGAAAGGV 85
              ::|| ||| |  : |  ||::|||   :  |       :|    || |: :|||||   :
Db         16 YVAALAAAGGAFAAGTVLGWTSPAETEIVDRGEGYDFPVDKDQFSWVGSAMTLGAACVCI 75

Qy         86 LGGWLVDRAGRKLSLLLCSVPFVAGFAVITAAQDVWMLLGGRLLTGLACGVASLVAPVYI 145
                |:|::  ||| ::|   :||: |: ::  | :| ||   | : |:| |   : ||:|
Db         76 PIGFLINMIGRKWTMLFLVLPFILGWTMLIWAVNVSMLYASRFILGIAGGAFCVTAPMYT 135

Qy        146 SEIAYPAVRGLLGSCVQLMVVVGILLAYLAGWVLEWRWLAVLGCVPPSLMLLLMCFMPET 205
               |||   :|| |||  |||: :|||  |  |  ::  ||:::  : | :   :  ||||:
Db        136 GEIAQKEIRGTLGSFFQLMITIGILFVYAVGAGVKIFWLSIICGILPLIFGAIFFFMPES 195

Qy        206 PRFLLTQHRRQEAMAALRFLWGSEQGWE---------DPPIGAEQSFHLALLRQPGIYKP 256
              | :|::: | : |: ::::| | |  :|         |    | :    | | :|   |
Db        196 PTYLVSKDRSENAIKSIQWLRGKEYDYEPELAELRETDRETKANKVNVWAALNRPVTRKA 255

Qy        257 FIIGVSLMAFQQLSGVNAVMFYAETIFEEAKFK-DSSLASVVVGVIQVLFTAVAALIMDR 315
                | : || |||: |:|||:|||  || ||    ::  |::::|::||: | |: |::|:
Db        256 LAISMGLMFFQQVCGINAVIFYASRIFLEANTGIEAEWATILIGIMQVVATFVSTLVVDK 315

Qy        316 AGRRLLLVLSGVVMVFSTSAFGAYFKLTQGGPGNSSHVAISAPVSAQPVDAS--VGLAWL 373
               |||:||: ||: |  ||:| | || |                   |  ||:  | | ||
Db        316 LGRRILLLASGISMAISTTAIGVYFFL-------------------QKQDAAQVVSLGWL 356

Qy        374 AVGSMCLFIAGFAVGWGPIPWLLMSEIFPLHVKGVATGICVLTNWLMAFLVTKEFSSLME 433
               | |:||||  |::|:||:|||:| |:|   :|| |  :   :|||:||:||| | :| :
Db        357 PVASLCLFIIMFSIGYGPVPWLMMGELFATDIKGFAGSLAGTSNWLLAFVVTKTFVNLND 416

Qy        434 VLRPYGAFWLASAFCIFSVLFTLFCVPETKGKTLEQITAHFEG 476
               |   | ||| :   :  |:|  | |||||||:| :|     |
Db        417 GLGIGGTFWLFAGLTVVGVIFVYFAVPETKGKSLNEIQQELAG 459

RESULT 15
Q9V610
ID   Q9V610      PRELIMINARY;      PRT;   433 AA.
AC   Q9V610;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   CG8234 protein.
GN   CG8234.
OS   Drosophila melanogaster (Fruit fly).
```

```
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
DR   EMBL; AE003825; AAF58630.1; -.
DR   FlyBase; FBgn0033644; CG8234.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0005351; F:sugar porter activity; IEA.
DR   GO; GO:0005215; F:transporter activity; IEA.
DR   GO; GO:0008643; P:carbohydrate transport; IEA.
DR   InterPro; IPR007114; MFS.
DR   InterPro; IPR005828; Sub_transporter.
DR   InterPro; IPR003663; Sugar_transpt.
DR   InterPro; IPR005829; Sug_transporter.
DR   Pfam; PF00083; sugar_tr; 1.
DR   PRINTS; PR00171; SUGRTRNSPORT.
DR   TIGRFAMs; TIGR00879; SP; 1.
DR   PROSITE; PS50850; MFS; 1.
DR   PROSITE; PS00216; SUGAR_TRANSPORT_1; 2.
DR   PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.
KW   Sugar transport; Transmembrane; Transport.
SQ   SEQUENCE   433 AA;  47357 MW;  63F0E13046A769DD CRC64;

  Query Match             31.1%;  Score 764.5;  DB 5;  Length 433;
  Best Local Similarity   38.2%;  Pred. No. 1.3e-43;
  Matches  161;  Conservative   74;  Mismatches  156;  Indels   31;  Gaps    5;

Qy         68 AASWFGAVVTLGAAAGGVLGGWLVDRAGRKLSLLLCSVPFVAGFAVITAAQDVWMLLGGR 127
              | || | :: | | |||: || |::  ||: ::|  :|||:    :|  | :| |:| ||
Db         15 AGSWVGGIMPLAALAGGITGGPLIEYLGRRSTILATAVPFIVSSLLIACAVNVIMILCGR 74

Qy        128 LLTGLACGVASLVAPVYISEIAYPAVRGLLGSCVQLMVVVGILLAYLAGWVLEWRWLAVL 187
               |||   |:|||  |||: |   | ||| ||     :  :|||: |:||  : |  || |
Db         75 FLTGFCVGIASLSLPVYLGETLQPEVRGTLGLLPTALGNIGILVCYVAGSFMNWSMLAFL 134

Qy        188 GCVPPSLMLLLMCFMPETPRFLLTQHRRQEAMAALRFLWGSE-----------QGWEDPP 236
              |   |   |:||  :|||||: : : : : |  ||::| | |           |   |
Db        135 GAALPVPFLILMIIIPETPRWFVNRGQEERARKALKWLRGKEADVEPELKELMQSQADAD 194

Qy        237 IGAEQSFHLALLRQPGIYKPFIIGVSLMAFQQLSGVNAVMFYAETIFEEA-KFKDSSLAS 295
                | |:  | | ::  : ||  | : || ||| ||:|||:||   ||::|    ||:|::
Db        195 RQATQNTCLELFKRNNL-KPLSISLGLMFFQQFSGINAVIFYTVQIFKDAGSTIDSNLST 253

Qy        296 VVVGVIQVLFTAVAALIMDRAGRRLLLVLSGVVMVFSTSAFGAYFKLTQGGPGNSSHVAI 355
              ::|||:    | :  :::|| ||::|| :| : |: : |  | :|     || : ||
Db        254 IIVGVVNFFATFMGIILIDRLGRKILLYVSDIAMIVTLSILGGFFYCKAHGP-DVSH--- 309

Qy        356 SAPVSAQPVDASVGLAWLAVGSMCLFIAGFAVGWGPIPWLLMSEIFPLHVKGVATGICVL 415
                            | || :    ::| ||::|:||||||:| || |  ::| |  :
Db        310 --------------LGWLPLTCFVIYILGFSLGFGPIPWLMMGEILPAKIRGPAASVVTA 355

Qy        416 TNWLMAFLVTKEFSSLMEVLRPYGAFWLASAFCIFSVLFTLFCVPETKGKTLEQITAHFE 475
               ||   |:||| |  |   :  :|||||  | ||  : | :  ||||:||:||:|
Db        356 FNWFCTFVVTKTFQDLTVAMGAHGAFWLFGAICIVGLFFVIIFVPETRGKSLEEIERKMM 415

Qy        476 GR 477
              ||
Db        416 GR 417


Search completed: September 27, 2004, 19:29:18
Job time : 123 secs
```

```
                    GenCore version 5.1.6
           Copyright (c) 1993 - 2004  Compugen Ltd.

OM protein - protein search, using sw model

Run on:          September 27, 2004, 19:29:30 ; Search time 147 Seconds
                                          (without alignments)
                                          1043.425 Million cell updates/sec

Title:           US-09-886-954A-1
Perfect score:   2457
Sequence:        1 MTPEDPEETQPLLGPPGGSA..........CVPETKGKTLEQITAHFEGR 477

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        1349238 seqs, 321558718 residues

Total number of hits satisfying chosen parameters:      1349238

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Published_Applications_AA:*
                1:  /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
                2:  /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
                3:  /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
                4:  /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
                5:  /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
                6:  /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
                7:  /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
                8:  /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
                9:  /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
                10:  /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
                11:  /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
                12:  /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
                13:  /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
                14:  /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
                15:  /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
                16:  /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
                17:  /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
                18:  /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                               SUMMARIES
                       %
Result               Query
  No.      Score     Match Length DB  ID                            Description
-------------------------------------------------------------------------------
     1      2457     100.0    477 14  US-10-168-651-1               Sequence 1, Appli
     2      2450      99.7    477 12  US-09-886-954-1               Sequence 1, Appli
     3      1491      60.7    326 15  ~~US-10-264-237-2093~~        Sequence 2093, Ap
     4      1135      46.2    262 12  US-10-169-395-9               Sequence 9, Appli
     5       948      38.6    507 14  US-10-157-031-359             Sequence 359, App
     6       948      38.6    507 16  US-10-755-889-110             Sequence 110, App
     7       723      29.4    501 13  US-10-051-909-38              Sequence 38, Appl
     8       716      29.1    484 12  US-10-424-599-275264          Sequence 275264,
     9       702      28.6    501 16  US-10-437-963-163544          Sequence 163544,
    10       702      28.6    523 12  US-10-425-114-63431           Sequence 63431, A
    11     701.5      28.6    509 12  US-10-425-114-59933           Sequence 59933, A
    12     658.5      26.8    553 13  US-10-051-909-36              Sequence 36, Appl
    13       639      26.0    481 15  US-10-310-154-725             Sequence 725, App
    14       615      25.0    488 12  US-09-794-822-11              Sequence 11, Appl
    15       615      25.0    488 13  US-10-094-059-4               Sequence 4, Appli
    16       615      25.0    488 14  US-10-170-528-5               Sequence 5, Appli
    17       615      25.0    488 14  US-10-162-012-46              Sequence 46, Appl
    18       615      25.0    488 14  US-10-062-960B-4              Sequence 4, Appli
    19       615      25.0    488 14  US-10-144-624-4               Sequence 4, Appli
    20       615      25.0    488 15  US-10-162-102-46              Sequence 46, Appl
    21       615      25.0    488 16  US-10-038-854-170             Sequence 170, App
    22     613.5      25.0    487  9  US-09-795-693-27              Sequence 27, Appl
    23     613.5      25.0    487 13  US-10-095-139-14              Sequence 14, Appl
    24     613.5      25.0    487 14  US-10-156-239-27              Sequence 27, Appl
    25     613.5      25.0    487 14  US-10-199-485-27              Sequence 27, Appl
    26     613.5      25.0    487 15  US-10-391-399-45              Sequence 45, Appl
    27     609.5      24.8    502 16  US-10-437-963-166774          Sequence 166774,
    28       607      24.7    461 15  US-10-369-493-23371           Sequence 23371, A
    29     604.5      24.6    592 12  US-10-425-114-63791           Sequence 63791, A
    30     604.5      24.6    596 16  US-10-437-963-177463          Sequence 177463,
    31     604.5      24.6    601 12  US-10-425-114-63432           Sequence 63432, A
    32       603      24.5    472 13  US-10-024-623-29              Sequence 29, Appl
    33       603      24.5    472 14  US-10-154-419-79              Sequence 79, Appl
    34       603      24.5    472 14  US-10-146-733-74              Sequence 74, Appl
    35       603      24.5    472 15  US-10-369-493-23536           Sequence 23536, A
    36     598.5      24.4    465 12  US-10-282-122A-56366          Sequence 56366, A
    37       598      24.3    486  9  US-09-860-232A-7              Sequence 7, Appli
    38     593.5      24.2    525 12  US-10-425-114-63426           Sequence 63426, A
    39     591.5      24.1    464 12  US-10-282-122A-42918          Sequence 42918, A
    40     591.5      24.1    464 13  US-10-024-623-28              Sequence 28, Appl
    41     591.5      24.1    464 14  US-10-154-419-78              Sequence 78, Appl
    42     591.5      24.1    464 14  US-10-146-733-73              Sequence 73, Appl
    43     591.5      24.1    464 15  US-10-369-493-23552           Sequence 23552, A
    44     591.5      24.1    523 16  US-10-437-963-124519          Sequence 124519,
    45     588.5      24.0    535  9  US-09-795-693-20              Sequence 20, Appl

                              ALIGNMENTS

RESULT 1
US-10-168-651-1
; Sequence 1, Application US/10168651
; Publication No. US20030171275A1
; GENERAL INFORMATION:
;  APPLICANT: INCYTE GENOMICS, INC.
;  APPLICANT:  BAUGHN, Mariah R.
;  APPLICANT:  BURFORD, Neil
;  APPLICANT:  AU-YOUNG, Janice
;  APPLICANT:  LU, Dyung Aina M.
;  APPLICANT:  YANG, Junming
;  APPLICANT:  REDDY, Roopa
;  APPLICANT:  LAL, Preeti
;  APPLICANT:  HILLMAN, Jennifer L.
;  APPLICANT:  AZIMZAI, Yalda
;  APPLICANT:  YUE, Henry
;  APPLICANT:  NGUYEN, Danniel B.
;  APPLICANT:  YAO, Monique G.
;  APPLICANT:  GANDHI, Ameeni R.
;  APPLICANT:  TANG, Y. Tom
;  APPLICANT:  KHAN, Farrah A.
;  TITLE OF INVENTION: TRANSPORTERS AND ION CHANNELS
;  FILE REFERENCE: PI-0005 PCT
;  CURRENT APPLICATION NUMBER: US/10/168,651
;  CURRENT FILING DATE:  2002-06-21
;  PRIOR APPLICATION NUMBER: 60/172,000; 60/176,083; 60/177,332; 60/178,572; 60/179,758
;    60/181,625
;  PRIOR FILING DATE: 1999-12-23; 2000-01-14; 2000-01-21; 2000-01-28; 2000-02-02;
;    2000-02-10
;  NUMBER OF SEQ ID NOS: 54
;  SOFTWARE: PERL Program
; SEQ ID NO 1
;   LENGTH: 477
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. US20030171275A1 1416107CD1
```

```
US-10-168-651-1

  Query Match             100.0%;  Score 2457;  DB 14;  Length 477;
  Best Local Similarity   100.0%;  Pred. No. 2.2e-208;
  Matches  477;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MTPEDPEETQPLLGPPGGSAPRGRRVFLAAFAAALGPLSFGFALGYSSPAIPSLQRAAPP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MTPEDPEETQPLLGPPGGSAPRGRRVFLAAFAAALGPLSFGFALGYSSPAIPSLQRAAPP 60

Qy         61 APRLDDAAASWFGAVVTLGAAAGGVLGGWLVDRAGRKLSLLLCSVPFVAGFAVITAAQDV 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 APRLDDAAASWFGAVVTLGAAAGGVLGGWLVDRAGRKLSLLLCSVPFVAGFAVITAAQDV 120

Qy        121 WMLLGGRLLTGLACGVASLVAPVYISEIAYPAVRGLLGSCVQLMVVVGILLAYLAGWVLE 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 WMLLGGRLLTGLACGVASLVAPVYISEIAYPAVRGLLGSCVQLMVVVGILLAYLAGWVLE 180

Qy        181 WRWLAVLGCVPPSLMLLLMCFMPETPRFLLTQHRRQEAMAALRFLWGSEQGWEDPPIGAE 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 WRWLAVLGCVPPSLMLLLMCFMPETPRFLLTQHRRQEAMAALRFLWGSEQGWEDPPIGAE 240

Qy        241 QSFHLALLRQPGIYKPFIIGVSLMAFQQLSGVNAVMFYAETIFEEAKFKDSSLASVVVGV 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QSFHLALLRQPGIYKPFIIGVSLMAFQQLSGVNAVMFYAETIFEEAKFKDSSLASVVVGV 300

Qy        301 IQVLFTAVAALIMDRAGRRLLLVLSGVVMVFSTSAFGAYFKLTQGGPGNSSHVAISAPVS 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 IQVLFTAVAALIMDRAGRRLLLVLSGVVMVFSTSAFGAYFKLTQGGPGNSSHVAISAPVS 360

Qy        361 AQPVDASVGLAWLAVGSMCLFIAGFAVGWGPIPWLLMSEIFPLHVKGVATGICVLTNWLM 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AQPVDASVGLAWLAVGSMCLFIAGFAVGWGPIPWLLMSEIFPLHVKGVATGICVLTNWLM 420

Qy        421 AFLVTKEFSSLMEVLRPYGAFWLASAFCIFSVLFTLFCVPETKGKTLEQITAHFEGR 477
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AFLVTKEFSSLMEVLRPYGAFWLASAFCIFSVLFTLFCVPETKGKTLEQITAHFEGR 477


RESULT 2
US-09-886-954-1
; Sequence 1, Application US/09886954
; Publication No. US20020038464A1
; GENERAL INFORMATION:
;  APPLICANT: Charron, Maureen J.
;  APPLICANT:  Katz, Ellen B.
;  TITLE OF INVENTION: NOVEL GLUCOSE TRANSPORTER/SENSOR PROTEIN AND USES THEREOF
;  FILE REFERENCE: 96700/667
;  CURRENT APPLICATION NUMBER: US/09/886,954
;  CURRENT FILING DATE:  2001-06-21
;  NUMBER OF SEQ ID NOS: 1
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
;   LENGTH: 477
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-886-954-1

  Query Match             99.7%;  Score 2450;  DB 12;  Length 477;
  Best Local Similarity   99.8%;  Pred. No. 9.1e-208;
  Matches  476;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 MTPEDPEETQPLLGPPGGSAPRGRRVFLAAFAAALGPLSFGFALGYSSPAIPSLQRAAPP 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db          1 MTPEDPEETQPLLGPPGGSAPRGRRVFLAAFAAALGPLSFGFALGYSSPAIPSLGRAAPP 60

Qy         61 APRLDDAAASWFGAVVTLGAAAGGVLGGWLVDRAGRKLSLLLCSVPFVAGFAVITAAQDV 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 APRLDDAAASWFGAVVTLGAAAGGVLGGWLVDRAGRKLSLLLCSVPFVAGFAVITAAQDV 120

Qy        121 WMLLGGRLLTGLACGVASLVAPVYISEIAYPAVRGLLGSCVQLMVVVGILLAYLAGWVLE 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 WMLLGGRLLTGLACGVASLVAPVYISEIAYPAVRGLLGSCVQLMVVVGILLAYLAGWVLE 180
```

```
Qy        181 WRWLAVLGCVPPSLMLLLMCFMPETPRFLLTQHRRQEAMAALRFLWGSEQGWEDPPIGAE 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 WRWLAVLGCVPPSLMLLLMCFMPETPRFLLTQHRRQEAMAALRFLWGSEQGWEDPPIGAE 240

Qy        241 QSFHLALLRQPGIYKPFIIGVSLMAFQQLSGVNAVMFYAETIFEEAKFKDSSLASVVVGV 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QSFHLALLRQPGIYKPFIIGVSLMAFQQLSGVNAVMFYAETIFEEAKFKDSSLASVVVGV 300

Qy        301 IQVLFTAVAALIMDRAGRRLLLVLSGVVMVFSTSAFGAYFKLTQGGPGNSSHVAISAPVS 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 IQVLFTAVAALIMDRAGRRLLLVLSGVVMVFSTSAFGAYFKLTQGGPGNSSHVAISAPVS 360

Qy        361 AQPVDASVGLAWLAVGSMCLFIAGFAVGWGPIPWLLMSEIFPLHVKGVATGICVLTNWLM 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AQPVDASVGLAWLAVGSMCLFIAGFAVGWGPIPWLLMSEIFPLHVKGVATGICVLTNWLM 420

Qy        421 AFLVTKEFSSLMEVLRPYGAFWLASAFCIFSVLFTLFCVPETKGKTLEQITAHFEGR 477
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AFLVTKEFSSLMEVLRPYGAFWLASAFCIFSVLFTLFCVPETKGKTLEQITAHFEGR 477


RESULT 3
US-10-264-237-2093
; Sequence 2093, Application US/10264237
; Publication No. US20040009491A1
; GENERAL INFORMATION:
;  APPLICANT: Birse et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PA131P1
;  CURRENT APPLICATION NUMBER: US/10/264,237
;  CURRENT FILING DATE:  2002-10-04
;  PRIOR APPLICATION NUMBER: PCT/US01/16450
;  PRIOR FILING DATE: 2001-05-18
;  PRIOR APPLICATION NUMBER: US 60/205,515
;  PRIOR FILING DATE: 2000-05-19
;  NUMBER OF SEQ ID NOS: 2876
;  SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 2093
;   LENGTH: 326
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (76)
;   OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (87)
;   OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (249)
;   OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (253)
;   OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (258)
;   OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (261)
;   OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (316)
;   OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
US-10-264-237-2093

  Query Match             60.7%;  Score 1491;  DB 15;  Length 326;
```

```
  Best Local Similarity   96.3%;  Pred. No. 3.2e-123;
  Matches  289;  Conservative    1;  Mismatches   10;  Indels    0;  Gaps    0;

Qy        164 MVVVGILLAYLAGWVLEWRWLAVLGCVPPSLMLLLMCFMPETPRFLLTQHRRQEAMAALR 223
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MVVVGILLAYLAGWVLEWRWLAVLGCVPPSLMLLLMCFMPETPRFLLTQHRRQEAMAALR 60

Qy        224 FLWGSEQGWEDPPIGAEQSFHLALLRQPGIYKPFIIGVSLMAFQQLSGVNAVMFYAETIF 283
              ||||||||||||||| |||||||||| |||||||||||||||||||||||||||||||||
Db         61 FLWGSEQGWEDPPIGXEQSFHLALLRXPGIYKPFIIGVSLMAFQQLSGVNAVMFYAETIF 120

Qy        284 EEAKFKDSSLASVVVGVIQVLFTAVAALIMDRAGRRLLLVLSGVVMVFSTSAFGAYFKLT 343
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 EEAKFKDSSLASVVVGVIQVLFTAVAALIMDRAGRRLLLVLSGVVMVFSTSAFGAYFKLT 180

Qy        344 QGGPGNSSHVAISAPVSAQPVDASVGLAWLAVGSMCLFIAGFAVGWGPIPWLLMSEIFPL 403
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 QGGPGNSSHVAISAPVSAQPVDASVGLAWLAVGSMCLFIAGFAVGWGPIPWLLMSEIFPL 240

Qy        404 HVKGVATGICVLTNWLMAFLVTKEFSSLMEVLRPYGAFWLASAFCIFSVLFTLFCVPETK 463
              |||||||| ||| |||| || |||||||||||||||||||||||||||||||| |  : |
Db        241 HVKGVATGXCVLXNWLMXFLXTKEFSSLMEVLRPYGAFWLASAFCIFSVLFTLVCALKLK 300


RESULT 4
US-10-169-395-9
; Sequence 9, Application US/10169395
; Publication No. US20040034192A1
; GENERAL INFORMATION:
;  APPLICANT: KATO, Seishi
;  APPLICANT:  KIMURA, Tomoko
;  TITLE OF INVENTION: HUMAN PROTEINS HAVING HYDROPHOBIC DOMAINS AND DNAs ENCODING
;  TITLE OF INVENTION:  THESE PROTEINS
;  FILE REFERENCE: 01997.015100.US
;  CURRENT APPLICATION NUMBER: US/10/169,395
;  CURRENT FILING DATE:  2002-11-29
;  PRIOR APPLICATION NUMBER: JP 2000-585
;  PRIOR FILING DATE: 2000-01-06
;  PRIOR APPLICATION NUMBER: JP 2000-588
;  PRIOR FILING DATE: 2000-01-06
;  PRIOR APPLICATION NUMBER: JP 2000-2299
;  PRIOR FILING DATE: 2000-01-11
;  PRIOR APPLICATION NUMBER: JP 2000-26862
;  PRIOR FILING DATE: 2000-02-03
;  PRIOR APPLICATION NUMBER: JP 2000-58367
;  PRIOR FILING DATE: 2000-03-03
;  PRIOR APPLICATION NUMBER: PCT/JP00/09359
;  PRIOR FILING DATE: 2000-12-28
;  NUMBER OF SEQ ID NOS: 150
; SEQ ID NO 9
;   LENGTH: 262
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-169-395-9

  Query Match             46.2%;  Score 1135;  DB 12;  Length 262;
  Best Local Similarity   91.4%;  Pred. No. 6.9e-92;
  Matches  223;  Conservative    2;  Mismatches    7;  Indels   12;  Gaps    1;

Qy          1 MTPEDPEETQPLLGPPGGSAPRGRRVFLAAFAAALGPLSFGFALGYSSPAIPSLQRAAPP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MTPEDPEETQPLLGPPGGSAPRGRRVFLAAFAAALGPLSFGFALGYSSPAIPSLQRAAPP 60

Qy         61 APRLDDAAASWFGAVVTLGAAAGGVLGGWLVDRAGRKLSLLLCSVPFVAGFAVITAAQDV 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 APRLDDAAASWFGAVVTLGAAAGGVLGGWLVDRAGRKLSLLLCSVPFVAGFAVITAAQDV 120

Qy        121 WMLLGGRLLTGLACGVASLVAPVYISEIAYPAVRGLLGSCVQLMVVVGILLAYLAGWVLE 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 WMLLGGRLLTGLACGVASLVAPVYISEIAYPAVRGLLGSCVQLMVVVGILLAYLAGWVLE 180

Qy        181 WRWLAVLGCVPPSLMLLLMCFMPETPRFLLTQHRRQEAM------------AALRFLWGS 228
              |||||||||||||||||||||||||||||||||||||||              || |  :
Db        181 WRWLAVLGCVPPSLMLLLMCFMPETPRFLLTQHRRQEAAPGLVRCGHGVQHECLRRLLQA 240

Qy        229 EQGW 232
              : ||
Db        241 DPGW 244


RESULT 5
US-10-157-031-359
; Sequence 359, Application US/10157031
; Publication No. US20030108890A1
; GENERAL INFORMATION:
;  APPLICANT: Baranova, A. V.
;  APPLICANT:  Yankovsky, N. K.
;  APPLICANT:  Kozlov, A. P.
;  APPLICANT:  Lobashev, A. V.
;  APPLICANT:  Krukovskaya, L. L.
;  TITLE OF INVENTION: In silico screening for phenotype-associated expressed sequences
;  FILE REFERENCE: 2760-103
;  CURRENT APPLICATION NUMBER: US/10/157,031
;  CURRENT FILING DATE:  2002-05-30
;  NUMBER OF SEQ ID NOS: 415
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 359
;   LENGTH: 507
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-157-031-359

  Query Match             38.6%;  Score 948;  DB 14;  Length 507;
  Best Local Similarity   42.7%;  Pred. No. 5.5e-75;
  Matches  216;  Conservative   80;  Mismatches  166;  Indels   44;  Gaps   11;

Qy         10 QPLLG------------PPGGSAPRGR-------RVFLAAFAAALGPLSFGFALGYSSPA 50
              :||||            ||     | |       ||||| ||| ||  |||:|  |:||
Db          3 EPLLGAEGPDYDTFPEKPPPSPGDRARVGTLQNKRVFLATFAAVLGNFSFGYALVYTSPV 62

Qy         51 IPSLQRAAPPAPRLDDAAASWFGAVVTLGAAAGGVLGGWLVDRAGRKLSLLLCSVPFVAG 110
              ||:|:|:  |   |  : |||||:| ||||||||:    | |  |||||::  :||  ||
Db         63 IPALERSLDPDLHLTKSQASWFGSVFTLGAAAGGLSAMILNDLLGRKLSIMFSAVPSAAG 122

Qy        111 FAVITAAQDVWMLLGGRLLTGLACGVASLVAPVYISEIAYPAVRGLLGSCVQLMVVVGIL 170
              :|::  |  :|||| || ||| | |: :   |||:|||| | ||| ||:  ||| | | |
Db        123 YALMAGAHGLWMLLLGRTLTGFAGGLTAACIPVYVSEIAPPGVRGALGATPQLMAVFGSL 182

Qy        171 LAYLAGWVLEWRWLAVLGCVPPSLMLLLMCFMPETPRFLLTQHRRQEAMAALRFLWGSEQ 230
                |  | :| |||||| |  |  :|:||: ||| :|||||:: | :||: || :| |::
Db        183 SLYALGLLLPWRWLAVAGEAPVLIMILLLSFMPNSPRFLLSRGRDEEALRALAWLRGTDV 242

Qy        231 G--WEDPPIG---AEQSFHL--ALLRQPGIYKPFIIGVSLMAFQQLSGVNAVMFYAETIF 283
                 ||   |      ||  :  |  | | : :|  : : :   |||:|:  :: | ::||
Db        243 DVHWEFEQIQDNVRRQSSRVSWAEARAPHVCRPITVALLMRLLQQLTGITPILVYLQSIF 302

Qy        284 EEAKF----KDSSLASVVVGVIQVLFTAVAALIMDRAGRRLLLVLSGVVMVFSTSAFGAY 339
              :        ||    : :|| :::|   :||| |||::|| :|  :|  :    | |
Db        303 DSTAVLLPPKDD---AAIVGAVRLLSVLIAALTMDLAGRKVLLFVSAAIMFAANLTLGLY 359

Qy        340 FKLTQGGPGNSSHVAISAPVS------AQPVDASVG-LAWLAVGSMCLFIAGFAVGWGPI 392
                    ||   |  : :   |      |||: |  | |  : : :  ||| |:|||||||
Db        360 IHF---GPRPLSPNSTAGLESESWGDLAQPLAAPAGYLTLVPLLATMLFIMGYAVGWGPI 416

Qy        393 PWLLMSEIFPLHVKGVATGICVLTNWLMAFLVTKEFSSLMEVLRPYGAFWLASAFCIFSV 452
               ||||||: ||  :|||:|:||| :|| ||::|| |  ::        |:  :| |: |:
Db        417 TWLLMSEVLPLRARGVASGLCVLASWLTAFVLTKSFLPVVSTFGLQVPFFFFAAICLVSL 476

Qy        453 LFTLFCVPETKGKTLEQITAHFE-GR 477
              :||  |||||||::|||| : |  ||
Db        477 VFTGCCVPETKGRSLEQIESFFRTGR 502


RESULT 6
US-10-755-889-110
; Sequence 110, Application US/10755889
```

```
; Publication No. US20040171823A1
; GENERAL INFORMATION:
;  APPLICANT: Bristol-Myers Squibb Company
;  TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ASSOCIATED WITH THE NF-kB
;  TITLE OF INVENTION:  PATHWAY
;  FILE REFERENCE: D0284 NP
;  CURRENT APPLICATION NUMBER: US/10/755,889
;  CURRENT FILING DATE:  2004-01-13
;  PRIOR APPLICATION NUMBER: U.S. 60/440,068
;  PRIOR FILING DATE: 2003-01-14
;  PRIOR APPLICATION NUMBER: U.S. 60/469,757
;  PRIOR FILING DATE: 2003-05-12
;  NUMBER OF SEQ ID NOS: 823
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 110
;   LENGTH: 507
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-755-889-110

  Query Match              38.6%;  Score 948;  DB 16;  Length 507;
  Best Local Similarity    42.7%;  Pred. No. 5.5e-75;
  Matches  216;  Conservative   80;  Mismatches  166;  Indels   44;  Gaps   11;

Qy           10 QPLLG------------PPGGSAPRGR-------RVFLAAFAAALGPLSFGFALGYSSPA 50
                :||||            ||     | |       ||||| ||| ||  |||:|| |:||
Db            3 EPLLGAEGPDYDTFPEKPPPSPGDRARVGTLQNKRVFLATFAAVLGNFSFGYALVYTSPV 62

Qy           51 IPSLQRAAPPAPRLDDAAASWFGAVVTLGAAAGGVLGGWLVDRAGRKLSLLLCSVPFVAG 110
                ||:|:|:  |   |  : |||||:| ||||||||:    | |  |||||::  :||  ||
Db           63 IPALERSLDPDLHLTKSQASWFGSVFTLGAAAGGLSAMILNDLLGRKLSIMFSAVPSAAG 122

Qy          111 FAVITAAQDVWMLLGGRLLTGLACGVASLVAPVYISEIAYPAVRGLLGSCVQLMVVVGIL 170
                :|::  |  :|||| || ||| | |: :   |||:|||| | ||| ||:  ||| | | |
Db          123 YALMAGAHGLWMLLLGRTLTGFAGGLTAACIPVYVSEIAPPGVRGALGATPQLMAVFGSL 182

Qy          171 LAYLAGWVLEWRWLAVLGCVPPSLMLLLMCFMPETPRFLLTQHRRQEAMAALRFLWGSEQ 230
                  |  | :| |||||| |  |  :|:||: ||| :|||||:: | :||: || :| |::
Db          183 SLYALGLLLPWRWLAVAGEAPVLIMILLLSFMPNSPRFLLSRGRDEEALRALAWLRGTDV 242

Qy          231 G--WEDPPIG---AEQSFHL--ALLRQPGIYKPFIIGVSLMAFQQLSGVNAVMFYAETIF 283
                   ||   |      ||  :  |  | | : :|  : : :   |||:|:  :: | ::||
Db          243 DVHWEFEQIQDNVRRQSSRVSWAEARAPHVCRPITVALLMRLLQQLTGITPILVYLQSIF 302

Qy          284 EEAKF----KDSSLASVVVGVIQVLFTAVAALIMDRAGRRLLLVLSGVVMVFSTSAFGAY 339
                :        ||    : :|| :::|   :||| || |||::|| :|  :|  :    | |
Db          303 DSTAVLLPPKDD---AAIVGAVRLLSVLIAALTMDLAGRKVLLFVSAAIMFAANLTLGLY 359

Qy          340 FKLTQGGPGNSSHVAISAPVS------AQPVDASVG-LAWLAVGSMCLFIAGFAVGWGPI 392
                      ||   |  : :   |      |||: |  | |  : : :  ||| |:|||||||
Db          360 IHF---GPRPLSPNSTAGLESESWGDLAQPLAAPAGYLTLVPLLATMLFIMGYAVGWGPI 416

Qy          393 PWLLMSEIFPLHVKGVATGICVLTNWLMAFLVTKEFSSLMEVLRPYGAFWLASAFCIFSV 452
                 ||||||: ||  :|||:|:||| :|| ||::|| |  ::        |:  :| |: |:
Db          417 TWLLMSEVLPLRARGVASGLCVLASWLTAFVLTKSFLPVVSTFGLQVPFFFFAAICLVSL 476

Qy          453 LFTLFCVPETKGKTLEQITAHFE-GR 477
                :||  |||||||::|||| : |  ||
Db          477 VFTGCCVPETKGRSLEQIESFFRMGR 502

RESULT 7
US-10-051-909-38
; Sequence 38, Application US/10051909
; Publication No. US20020199217A1
; GENERAL INFORMATION:
;  APPLICANT: Allen, Steve
;  APPLICANT:  Helentjaris, Tim
;  APPLICANT:  Hitz, Bill
;  APPLICANT:  Kinney, Tony
;  APPLICANT:  Tingey, Scott
;  TITLE OF INVENTION: Plant Sugar Transport Proteins
;  FILE REFERENCE: BB1163 US CIP
```

```
;  CURRENT APPLICATION NUMBER: US/10/051,909
;  CURRENT FILING DATE:  2002-01-17
;  PRIOR APPLICATION NUMBER: 60/083,044
;  PRIOR FILING DATE: April 24, 1998
;  NUMBER OF SEQ ID NOS: 38
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 38
;   LENGTH: 501
;   TYPE: PRT
;   ORGANISM: Oryza sativa
US-10-051-909-38

  Query Match              29.4%;  Score 723;  DB 13;  Length 501;
  Best Local Similarity    35.6%;  Pred. No. 4.2e-55;
  Matches  173;  Conservative   89;  Mismatches  172;  Indels   52;  Gaps    8;

Qy           17 GGSAPRGRRVFLAAFA--------------AALGPLSFGFALGYSSPAIPSLQRAAPPAP 62
                ||    | |:  :|::               ||||: |||  |:|||     | |
Db           39 GGGGGMGSRLGSSAYSLRDSSVSAVLCTLIVALGPIQFGFTCGFSSPT----QDAIISDL 94

Qy           63 RLDDAAASWFGAVVTLGAAAGGVLGGWLVDRAGRKLSLLLCSVPFVAGFAVITAAQDVWM 122
                 |  :  | ||::  :||  | :  | : :  ||| ||:: ::| : |:  |: |:|
Db           95 GLTLSEFSLFGSLSNVGAMVGAIASGQIAEYIGRKGSLMIAAIPNIIGWLAISFAKDSSF 154

Qy          123 LLGGRLLTGLACGVASLVAPVYISEIAYPAVRGLLGSCVQLMVVVGILLAYLAGWVLEWR 182
                |  |||| |   || | | ||||:|||   :|| |||  || | :||||||| |  : ||
Db          155 LFMGRLLEGFGVGVISYVVPVYIAEIAPQTMRGALGSVNQLSVTIGILLAYLLGMFVPWR 214

Qy          183 WLAVLGCVPPSLMLLLMCFMPETPRFLLTQHRRQEAMAALRFLWGSEQGWEDPPIGAE-- 240
                 |:||| :| |:::  : |:||:||:|    : ::  ::|: | | |       |  |
Db          215 ILSVLGILPCSILIPGLFFIPESPRWLAKMGKMEDFESSLQVLRGFE-----TDIAVEVN 269

Qy          241 ------------QSFHLALLRQPGIYKPFIIGVSLMAFQQLSGVNAVMFYAETIFEEAKF 288
                             :   | ::|     | ::|: |:  ||||||| ::||| :||: |
Db          270 EIKRSVQSSRRRTTIRFADIKQKRYSVPLMVGIGLLVLQQLSGVNGILFYAASIFKAAGL 329

Qy          289 KDSSLASVVVGVIQVLFTAVAALIMDRAGRRLLLVLSGVVMVFSTSAFGAYFKLTQGGPG 348
                 :|:||:  :||:||: | |   : |:|||||||::|   |  :       |   :
Db          330 TNSNLATFGLGVVQVVATGVTTWLTDKAGRRLLLIISTTGMTITLVVVSVSF-FVKDNIT 388

Qy          349 NSSHVAISAPVSAQPVDASVGLAWLAVGSMCLFIAGFAVGWGPIPWLLMSEIFPLHVKGV 408
                | ||:             || :: |::  :  |:  |::| | |||::|||| |:::| :
Db          389 NGSHL------------YSV-MSMLSLVGLVAFVISFSLGLGAIPWIIMSEILPVNIKSL 435

Qy          409 ATGICVLTNWLMAFLVTKEFSSLMEVLRPYGAFWLASAFCIFSVLFTLFCVPETKGKTLE 468
                |  :  | ||| |:|:|   :|||      | | : :| |  :::|    ||||||:|||
Db          436 AGSVATLANWLTAWLITMT-ASLMLSWSNGGTFAIYAAVCAGTLVFVCLWVPETKGRTLE 494

Qy          469 QITAHF 474
                :|   |
Db          495 EIAFSF 500

RESULT 8
US-10-424-599-275264
; Sequence 275264, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa Thomas J
;  APPLICANT:  Kovalic David K
;  APPLICANT:  Zhou Yihua
;  APPLICANT:  Cao Yongwei
;  TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53223)B
;  CURRENT APPLICATION NUMBER: US/10/424,599
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 275264
;   LENGTH: 484
;   TYPE: PRT
;   ORGANISM: Glycine max
;   FEATURE:
```

```
;   OTHER INFORMATION: Clone ID: PAT_MRT3847_90584C.1.pep
US-10-424-599-275264

  Query Match             29.1%;  Score 716;  DB 12;  Length 484;
  Best Local Similarity   35.6%;  Pred. No. 1.7e-54;
  Matches  164;  Conservative   89;  Mismatches  174;  Indels   34;  Gaps    7;

Qy          26 VFLAAFAAALGPLSFGFALGYSSPAIPSLQRAAPPAPRLDDAAASWFGAVVTLGAAAGGV 85
               |       ||||: |||  |||||   :: |       |  :  |:||::  :||  | :
Db          45 VLFCVLIVALGPIQFGFTCGYSSPTQGAIVRDL----NLSISEFSFFGSLSNVGAMVGAI 100

Qy          86 LGGWLVDRAGRKLSLLLCSVPFVAGFAVITAAQDVWMLLGGRLLTGLACGVASLVAPVYI 145
                 | : :  ||| ||:: ::| : |:  |: |:|   |  |||| |   |: | | ||||
Db         101 ASGQIAEYIGRKGSLMIAAIPNIIGWLAISFAKDSSFLYMGRLLEGFGVGIISYVVPVYI 160

Qy         146 SEIAYPAVRGLLGSCVQLMVVVGILLAYLAGWVLEWRWLAVLGCVPPSLMLLLMCFMPET 205
               :|||   :|| |||  || | :||:|||| |  : || ||:|| :| ::::  : |:||:
Db         161 AEIAPQNLRGGLGSVNQLSVTIGIMLAYLLGLFVNWRVLAILGILPCTVLIPGLFFIPES 220

Qy         206 PRFLLTQHRRQEAMAALRFLWGSEQGWEDPPI------------GAEQSFHLALLRQPGI 253
               ||:|       |   :|: | | :    |  :            |   :   | |::
Db         221 PRWLAKMGMIDEFETSLQVLRGFD---TDISVEVHEIKRSVASTGKRAAIRFADLKRKRY 277

Qy         254 YKPFIIGVSLMAFQQLSGVNAVMFYAETIFEEAKFKDSSLASVVVGVIQVLFTAVAALIM 313
               : | ::|: |:  |||||:| ::||: |||  |    |  |:| :| :||: | ::  ::
Db         278 WFPLMVGIGLLVLQQLSGINGILFYSTTIFANAGISSSEAATVGLGAVQVIATGISTWLV 337

Qy         314 DRAGRRLLLVLSGVVMVFSTSAFGAYFKLTQGGPGNSSHVAISAPVSAQPVDASVGLAWL 373
               |::||||||::|  ||  |       | | :|     ||:      |   : : |||  :
Db         338 DKSGRRLLLIISSSVMTVSLLIVSIAFYL-EGVVSEDSHL-----FSILGIVSIVGLVAM 391

Qy         374 AVGSMCLFIAGFAVGWGPIPWLLMSEIFPLHVKGVATGICVLTNWLMAFLVTKEFSSLME 433
                :        ||::| ||||||:|||| |:::||:|  |  : |||::: :|   : |:
Db         392 VI--------GFSLGLGPIPWLIMSEILPVNIKGLAGSIATMGNWLISWGITMTANLLLN 443

Qy         434 VLRPYGAFWLASAFCIFSVLFTLFCVPETKGKTLEQITAHF 474
                    | | : :    |:: |    ||||||:|||:|   |
Db         444 -WSSGGTFTIYTVVAAFTIAFIAMWVPETKGRTLEEIQFSF 483

RESULT 9
US-10-437-963-163544
; Sequence 163544, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  APPLICANT:  Wu, Wei
;  APPLICANT:  Boukharov, Andrey A.
;  APPLICANT:  Barbazuk, Brad
;  APPLICANT:  Li, Ping
;  TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53221)B
;  CURRENT APPLICATION NUMBER: US/10/437,963
;  CURRENT FILING DATE:  2003-05-14
;  NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 163544
;   LENGTH: 501
;   TYPE: PRT
;   ORGANISM: Oryza sativa
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_62530C.1.pep
US-10-437-963-163544

  Query Match             28.6%;  Score 702;  DB 16;  Length 501;
  Best Local Similarity   35.9%;  Pred. No. 3e-53;
  Matches  166;  Conservative   80;  Mismatches  178;  Indels   38;  Gaps    6;

Qy          27 FLAAFAAALGPLSFGFALGYSSPAIPSLQRAAPPAPRLDDAAASWFGAVVTLGAAAGGVL 86
               ||     ||||: |||  |:|||   :: |       |  :  | ||::  :||  | :

Db          63 FLCTLIVALGPIQFGFTGGFSSPTQDAIIRDLD----LTLSEFSVFGSLSNVGAMVGAIA 118

Qy          87 GGWLVDRAGRKLSLLLCSVPFVAGFAVITAAQDVWMLLGGRLLTGLACGVASLVAPVYIS 146
                | : :  ||| ||:: ::| : |:  |: |:|   |  |||| |   || |   ||||:
Db         119 SGQMAEYIGRKGSLMIAAIPNIIGWLAISFAKDSSFLYMGRLLEGFGVGVISYTVPVYIA 178

Qy         147 EIAYPAVRGLLGSCVQLMVVVGILLAYLAGWVLEWRWLAVLGCVPPSLMLLLMCFMPETP 206
               ||:   :|| |||  || | |||||||| |  : || |||:| :| ::::  : |:||:|
Db         179 EISPQNMRGALGSVNQLSVTVGILLAYLLGMFVPWRLLAVIGILPCTVLIPGLFFIPESP 238

Qy         207 RFLLTQHRRQEAMAALRFLWGSEQGWEDPPIGAE--------------QSFHLALLRQPG 252
               |:|   :   :   :|: | | |       | ||               :     | |
Db         239 RWLAKMNMMDDFETSLQVLRGFE-----TDISAEVNDIKRAVASANKRTTIRFQELNQKK 293

Qy         253 IYKPFIIGVSLMAFQQLSGVNAVMFYAETIFEEAKFKDSSLASVVVGVIQVLFTAVAALI 312
                  | |:|: |:  |||||:| ::||| :||: |   :| ||:  :| |||| | |   :
Db         294 YRTPLILGIGLLVLQQLSGINGILFYAGSIFKAAGLTNSDLATCALGAIQVLATGVTTWL 353

Qy         313 MDRAGRRLLLVLSGVVMVFSTSAFGAYFKLTQGGPGNSSHVAISAPVSAQPVDASVGLAW 372
               :||||||:||::|   |  |  |    | | :      ||:  :             |:
Db         354 LDRAGRRILLIISSAGMTLSLLAVAVVFFL-KDSISQDSHMYYT-------------LSM 399

Qy         373 LAVGSMCLFIAGFAVGWGPIPWLLMSEIFPLHVKGVATGICVLTNWLMAFLVTKEFSSLM 432
               ::: ::  |:  |: | | |||::|||| |: :| :|     | ||| :| :|   ::||
Db         400 ISLVALVAFVIAFSFGMGAIPWIIMSEILPVSIKSLAGSFATLANWLTSFGITMT-ANLM 458

Qy         433 EVLRPYGAFWLASAFCIFSVLFTLFCVPETKGKTLEQITAHF 474
                     | |        |:::| :  ||||||:|||:|   |
Db         459 LSWSAGGTFVSYMVVSAFTLVFVILWVPETKGRTLEEIQWSF 500

RESULT 10
US-10-425-114-63431
; Sequence 63431, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
;  APPLICANT: Liu, Jingdong
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Screen, Steven E
;  APPLICANT:  Tabaska, Jack E
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53313)B
;  CURRENT APPLICATION NUMBER: US/10/425,114
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 63431
;   LENGTH: 523
;   TYPE: PRT
;   ORGANISM: Oryza sativa
;   FEATURE:
;   OTHER INFORMATION: Clone ID: LIB4371-017-F11_FLI.pep
US-10-425-114-63431

  Query Match             28.6%;  Score 702;  DB 12;  Length 523;
  Best Local Similarity   35.9%;  Pred. No. 3.2e-53;
  Matches  166;  Conservative   80;  Mismatches  178;  Indels   38;  Gaps    6;

Qy          27 FLAAFAAALGPLSFGFALGYSSPAIPSLQRAAPPAPRLDDAAASWFGAVVTLGAAAGGVL 86
               ||     ||||: |||  |:|||   :: |       |  :  | ||::  :||  | :
Db          85 FLCTLIVALGPIQFGFTGGFSSPTQDAIIRDLD----LTLSEFSVFGSLSNVGAMVGAIA 140

Qy          87 GGWLVDRAGRKLSLLLCSVPFVAGFAVITAAQDVWMLLGGRLLTGLACGVASLVAPVYIS 146
                | : :  ||| ||:: ::| : |:  |: |:|   |  |||| |   || |   ||||:
Db         141 SGQMAEYIGRKGSLMIAAIPNIIGWLAISFAKDSSFLYMGRLLEGFGVGVISYTVPVYIA 200

Qy         147 EIAYPAVRGLLGSCVQLMVVVGILLAYLAGWVLEWRWLAVLGCVPPSLMLLLMCFMPETP 206
               ||:   :|| |||  || | |||||||| |  : || |||:| :| ::::  : |:||:|
Db         201 EISPQNMRGALGSVNQLSVTVGILLAYLLGMFVPWRLLAVIGILPCTVLIPGLFFIPESP 260

Qy         207 RFLLTQHRRQEAMAALRFLWGSEQGWEDPPIGAE--------------QSFHLALLRQPG 252
```

```
                |:|   :   :   :|: | | |       | ||               :     | |
Db          261 RWLAKMNMMDDFETSLQVLRGFE-----TDISAEVNDIKRAVASANKRTTIRFQELNQKK 315

Qy          253 IYKPFIIGVSLMAFQQLSGVNAVMFYAETIFEEAKFKDSSLASVVVGVIQVLFTAVAALI 312
                   | |:|: |:  |||||:| ::||| :||: |   :| ||:  :| |||| | |   :
Db          316 YRTPLILGIGLLVLQQLSGINGILFYAGSIFKAAGLTNSDLATCALGAIQVLATGVTTWL 375

Qy          313 MDRAGRRLLLVLSGVVMVFSTSAFGAYFKLTQGGPGNSSHVAISAPVSAQPVDASVGLAW 372
                :||||||:||::|   |  |  |    | | :      ||:  :             |:
Db          376 LDRAGRRILLIISSAGMTLSLLAVAVVFFL-KDSISQDSHMYYT-------------LSM 421

Qy          373 LAVGSMCLFIAGFAVGWGPIPWLLMSEIFPLHVKGVATGICVLTNWLMAFLVTKEFSSLM 432
                ::: ::  |:  |: | | |||::|||| |: :| :|     | ||| :| :|   ::||
Db          422 ISLVALVAFVIAFSFGMGAIPWIIMSEILPVSIKSLAGSFATLANWLTSFGITMT-ANLM 480

Qy          433 EVLRPYGAFWLASAFCIFSVLFTLFCVPETKGKTLEQITAHF 474
                      | |        |:::| :  ||||||:|||:|   |
Db          481 LSWSAGGTFVSYMVVSAFTLVFVILWVPETKGRTLEEIQWSF 522


RESULT 11
US-10-425-114-59933
; Sequence 59933, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
;  APPLICANT: Liu, Jingdong
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Screen, Steven E
;  APPLICANT:  Tabaska, Jack E
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53313)B
;  CURRENT APPLICATION NUMBER: US/10/425,114
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 59933
;   LENGTH: 509
;   TYPE: PRT
;   ORGANISM: Zea mays
;   FEATURE:
;   OTHER INFORMATION: Clone ID: LIB3689-227-G11_FLI.pep
US-10-425-114-59933

  Query Match            28.6%;  Score 701.5;  DB 12;  Length 509;
  Best Local Similarity  34.2%;  Pred. No. 3.4e-53;
  Matches  173;  Conservative   86;  Mismatches  182;  Indels   65;  Gaps    8;

Qy           10 QPLLG-------PPGGSAPRGR------RVFLAAFA--------------AALGPLSFGF 42
                :||:        || |     |      |:  :||:               ||||: |||
Db           27 KPLINTGSWYRMPPAGGVMGSRQSSLMERLGSSAFSLRDVAISATLCTLIVALGPIQFGF 86

Qy           43 ALGYSSPAIPSLQRAAPPAPRLDDAAASWFGAVVTLGAAAGGVLGGWLVDRAGRKLSLLL 102
                  |||||     | |      |  :  | ||::  :||  | :  | | :  ||| ||::
Db           87 TCGYSSPT----QDAIIADLGLSLSEFSLFGSLSNVGAMVGAISSGQLAEYIGRKGSLMI 142

Qy          103 CSVPFVAGFAVITAAQDVWMLLGGRLLTGLACGVASLVAPVYISEIAYPAVRGLLGSCVQ 162
                 ::| : |:  |: |:|   |  |||| |   || |   ||||:|||    || |||  |
Db          143 AAIPNIIGWLAISFAKDSSFLFMGRLLEGFGVGVISYTVPVYIAEIAPQDQRGALGSVNQ 202

Qy          163 LMVVVGILLAYLAGWVLEWRWLAVLGCVPPSLMLLLMCFMPETPRFLLTQHRRQEAMAAL 222
                | | :||||||| | :  || ||||| :| |:::  : |:||:||:|    : ::   :|
Db          203 LSVTIGILLAYLFGMFVPWRILAVLGILPCSILIPGLFFVPESPRWLAKMGKMEDFEYSL 262

Qy          223 RFLWGSEQGWEDPPIGAE--------------QSFHLALLRQPGIYKPFIIGVSLMAFQQ 268
                : | | :       | ||               :   | ::|     | :||: |:  ||
Db          263 QVLRGFQ-----TDITAEVNEIKRSLASSRRRTTIRFADIKQKRYSVPLVIGIGLLVLQQ 317

Qy          269 LSGVNAVMFYAETIFEEAKFKDSSLASVVVGVIQVLFTAVAALIMDRAGRRLLLVLSGVV 328
                ||||| ::||| :||: |   :|:||:  :| :||: | |   : |:|||||||::|
Db          318 LSGVNGILFYAASIFKAAGITNSNLATFGLGAVQVIATGVTTWLTDKAGRRLLLIISTTG 377

Qy          329 MVFSTSAFGAYFKLTQGGPGNSSHVAISAPVSAQPVDASVGLAWLAVGSMCLFIAGFAVG 388
                || :       | :       |   ::              :: |::  :  |:  |::|
Db          378 MVITLVIVSVSFFVKDNIAAGSHLYSV--------------MSMLSLAGLVAFVIAFSLG 423

Qy          389 WGPIPWLLMSEIFPLHVKGVATGICVLTNWLMAFLVTKEFSSLMEVLRPYGAFWLASAFC 448
                 | |||::|||| |:::| :|  :  | ||| |: :|   :|||      | | : :
Db          424 LGAIPWIIMSEILPVNIKSLAGSVATLANWLTAWAITMT-ASLMLNWSSGGTFAIYAVVS 482

Qy          449 IFSVLFTLFCVPETKGKTLEQITAHF 474
                  :::|    ||||||:|||:|   |
Db          483 TMALIFVCLWVPETKGRTLEEIAFSF 508


RESULT 12
US-10-051-909-36
; Sequence 36, Application US/10051909
; Publication No. US20020199217A1
; GENERAL INFORMATION:
;  APPLICANT: Allen, Steve
;  APPLICANT:  Helentjaris, Tim
;  APPLICANT:  Hitz, Bill
;  APPLICANT:  Kinney, Tony
;  APPLICANT:  Tingey, Scott
;  TITLE OF INVENTION: Plant Sugar Transport Proteins
;  FILE REFERENCE: BB1163 US CIP
;  CURRENT APPLICATION NUMBER: US/10/051,909
;  CURRENT FILING DATE:  2002-01-17
;  PRIOR APPLICATION NUMBER: 60/083,044
;  PRIOR FILING DATE: April 24, 1998
;  NUMBER OF SEQ ID NOS: 38
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 36
;   LENGTH: 553
;   TYPE: PRT
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: UNSURE
;   LOCATION: (528)
;   OTHER INFORMATION: Xaa = ANY AMINO ACID
US-10-051-909-36

  Query Match            26.8%;  Score 658.5;  DB 13;  Length 553;
  Best Local Similarity  34.0%;  Pred. No. 2.4e-49;
  Matches  159;  Conservative   77;  Mismatches  189;  Indels   43;  Gaps    6;

Qy           16 PGGSAPRGRR-----VFLAAFAAALGPLSFGFALGYSSPAIPSLQRAAPPAPRLDDAAAS 70
                || |:    |      ||     ||||: |||  |:|||   :: |       |  :  |
Db           84 PGTSSMAVLRESHVSAFLCTLIVALGPIQFGFTSGFSSPTQDAMVRDL----NLSISEFS 139

Qy           71 WFGAVVTLGAAAGGVLGGWLVDRAGRKLSLLLCSVPFVAGFAVITAAQDVWMLLGGRLLT 130
                 ||::  :|   | :  | : :  ||| ||:: ::| : |:  |: |:|   |  ||||
Db          140 AFGSLSNVGGMVGAIASGQMAEYIGRKGSLMIAAIPNIIGWLAISFAKDASFLYMGRLLE 199

Qy          131 GLACGVASLVAPVYISEIAYPAVRGLLGSCVQLMVVVGILLAYLAGWVLEWRWLAVLGCV 190
                |   |: |   ||||:||:   :|| |||  || |  || |||| |  : || |||:| :
Db          200 GFGVGIISYTVPVYIAEISPQNMRGALGSVNQLSVTFGIFLAYLLGMFIPWRLLAVIGAL 259

Qy          191 PPSLMLLLMCFMPETPRFLLTQHRRQEAMAALRFLWGSEQGWEDPPIGAE---------- 240
                | ::::  : |:||:||:|   :  ::   :|: | | |       |  |
Db          260 PCTMLIPGLFFIPESPRWLAKMNLTEDCETSLQVLRGFE-----TDITTEVNDIKRAVAS 314

Qy          241 ----QSFHLALLRQPGIYKPFIIGVSLMAFQQLSGVNAVMFYAETIFEEAKFKDSSLASV 296
                     :     | |     | ::|: |:  | |||:| |:||| :||: |   :| ||:
Db          315 SSKRTTISFQELNQKKYRTPLLLGIGLLVLQNLSGINGVLFYASSIFKAAGVTNSDLATC 374

Qy          297 VVGVIQVLFTAVAALIMDRAGRRLLLVLSGVVMVFSTSAFGAYFKLTQGGPGNSSHVAIS 356
                 :| |||| | |   ::||||||:||::|   |     |    | |      :|:   |
Db          375 SLGAIQVLATGVTTWLLDRAGRRILLIISTSGMTLCLLAVSVVFFLKDNISQDSNSYYI- 433

Qy          357 APVSAQPVDASVGLAWLAVGSMCLFIAGFAVGWGPIPWLLMSEIFPLHVKGVATGICVLT 416
                             |  :::  :  |:  |: | | ||||:|||| |: :| :   |  |
Db          434 -------------LTMISLVGIVSFVITFSFGMGAIPWLMMSEILPVSIKSLGGSIATLA 480
```

```
Qy        417 NWLMAFLVTKEFSSLMEVLRPYGAFWLASAFCIFSVLFTLFCVPETKG 464
              ||| :| :|   ::||      | |        |:::| :  ||||||
Db        481 NWLTSFAITMT-TNLMLTWSVGGTFLSYMVVSAFTIVFVVLWVPETKG 527

RESULT 13
US-10-310-154-725
; Sequence 725, Application US/10310154
; Publication No. US20030233670A1
; GENERAL INFORMATION:
;  APPLICANT: Edgerton, Michael D
;  APPLICANT:  Chomet, Paul S.
;  APPLICANT:  Adams, Thomas H
;  APPLICANT:  Ruff, Thomas G.
;  APPLICANT:  Agarwal, Ameeta K.
;  APPLICANT:  Ahrens, Jeffrey E.
;  APPLICANT:  Ball, James A.
;  APPLICANT:  Banu, G.
;  APPLICANT:  Bell, Erin
;  APPLICANT:  Boddupalli, Raghava
;  APPLICANT:  Deikman, Jill
;  APPLICANT:  Deng, Molian
;  APPLICANT:  Dong, Jinzhuo
;  APPLICANT:  Duff, Stephen M.
;  APPLICANT:  Galligan, Meghan M.
;  APPLICANT:  Hinchey, Brenda S.
;  APPLICANT:  Huang, Shihshieh
;  APPLICANT:  Johnson, G. Richard
;  APPLICANT:  Jung, Vincent
;  APPLICANT:  Kretzmer, Keith A
;  APPLICANT:  Laccetti, Lucille B.
;  APPLICANT:  Lai, Chao-Qiang
;  APPLICANT:  Lee, Gary
;  APPLICANT:  Lin, Jie-Yi
;  APPLICANT:  Liu, Jingdong
;  APPLICANT:  Lu, Bin
;  APPLICANT:  Luethy, Michael M.
;  APPLICANT:  Lund, Adrian
;  APPLICANT:  Madson, Linda L.
;  APPLICANT:  Malloy, Kathleen A.
;  APPLICANT:  McKiel, Christine L.
;  APPLICANT:  Miller, Philip W.
;  APPLICANT:  Padmavathi, Manchikanti
;  APPLICANT:  Parnell, Laurence D.
;  APPLICANT:  Start, William G.
;  APPLICANT:  Tennesen, Dan
;  APPLICANT:  Vidya, K.R.
;  APPLICANT:  Wang, Haiyun
;  APPLICANT:  Xin, Zhanguo
;  APPLICANT:  Xu, Nanfei
;  APPLICANT:  Yang, Chunzhi
;  APPLICANT:  Zeng, Xiaoping
;  APPLICANT:  Zhang, Qiang
;  APPLICANT:  Zhao, Yajuan
;  APPLICANT:  Zhou, Li
;  TITLE OF INVENTION: Gene Sequences and Uses Thereof in Plants
;  FILE REFERENCE: 38-15(52796)B
;  CURRENT APPLICATION NUMBER: US/10/310,154
;  CURRENT FILING DATE:  2002-12-04
;  PRIOR APPLICATION NUMBER: 60/337,358
;  PRIOR FILING DATE: 2001-12-04
;  NUMBER OF SEQ ID NOS: 736
; SEQ ID NO 725
;   LENGTH: 481
;   TYPE: PRT
;   ORGANISM: Glycine max
US-10-310-154-725

  Query Match             26.0%;  Score 639;  DB 15;  Length 481;
  Best Local Similarity   34.9%;  Pred. No. 1e-47;
  Matches  161;  Conservative   76;  Mismatches  182;  Indels   42;  Gaps    7;

Qy         26 VFLAAFAAALGPLSFGFALGYSSPAIPSLQRAAPPAPRLDDAAASWFGAVVTLGAAAGGV 85
```

```
              | |:   |  |  :||  :|||:|     | |      |  |  | ||::||:||  | :
Db         45 VLLSTLVAVCGSFTFGTCVGYSAPT----QAAIRADLNLSLAEFSMFGSLVTIGAMLGAI 100

Qy         86 LGGWLVDRAGRKLSLLLCSVPFVAGFAVITAAQDVWMLLGGRLLTGLACGVASLVAPVYI 145
                | : |  ||| :: : :   : |:  :  ::  : |  ||  ||   || | | ||||
Db        101 TSGRITDFIGRKGAMRISTGFCITGWIAVFFSKGSYSLDFGRFFTGYGIGVISYVVPVYI 160

Qy        146 SEIAYPAVRGLLGSCVQLMVVVGILLAYLAGWVLEWRWLAVLGCVPPSLMLLLMCFMPET 205
              :|||   :|| | :  ||::| |  :::| | |: || ||: | ||   :|: :||:||:
Db        161 AEIAPKNLRGGLATTNQLLIVTGGSVSFLLGSVINWRELALAGLVPCICLLVGLCFIPES 220

Qy        206 PRFLLTQHRRQEAMAALRFLWGSEQGWEDPP------IGAEQSF---HLALLRQPGIYKP 256
              ||:|    | :|   ||  | | :    |        |   ||     |  | |
Db        221 PRWLAKVGREKEFQLALSRLRGKDADISDEAAEILDYIETLQSLPKTKLLDLFQSKYVHS 280

Qy        257 FIIGVSLMAFQQLSGVNAVMFYAETIFEEAKFKDSSLASVVVGVIQVLFTAVAALIMDRA 316
               :||| ||| ||  |:| : ||   ||  |        ::    ||: || : |::||::
Db        281 VVIGVGLMACQQSVGINGIGFYTAEIFVAAGLSSGKAGTIAYACIQIPFTLLGAILMDKS 340

Qy        317 GRRLLLVLS--GVVMVFSTSAFGAYFKLTQGGPGNSSHVAISAPVSAQPVDASVGLAW-- 372
              ||| |:::|  |  :    :||  : |                       | |:   |
Db        341 GRRPLVMVSAAGTFLGCFVAAFAFFLK-----------------------DQSLLPEWVP 377

Qy        373 -LAVGSMCLFIAGFAVGWGPIPWLLMSEIFPLHVKGVATGICVLTNWLMAFLVTKEFSSL 431
               ||   : ::|| |::| | :||::||||||:|:|| |  : ||  || |::|:  |: |
Db        378 ILAFAGVLIYIAAFSIGLGSVPWVIMSEIFPIHLKGTAGSLVVLVAWLGAWVVSYTFNFL 437

Qy        432 MEVLRPYGAFWLASAFCIFSVLFTLFCVPETKGKTLEQITA 472
              |    | |  :| :   : ::||    ||||||||||:| |
Db        438 MSWSSP-GTLFLYAGCSLLTILFVAKLVPETKGKTLEEIQA 477

RESULT 14
US-09-794-822-11
; Sequence 11, Application US/09794822
; Publication No. US20030017585A1
; GENERAL INFORMATION:
;  APPLICANT: CuraGen Corporation
;  TITLE OF INVENTION: Novel Sodium/Solute Symporter-Like Protein and Nucleic
;  TITLE OF INVENTION:  Acids Encoding Same
;  FILE REFERENCE: 15966-687
;  CURRENT APPLICATION NUMBER: US/09/794,822
;  CURRENT FILING DATE:  2001-02-26
;  PRIOR APPLICATION NUMBER: 60/185,198
;  PRIOR FILING DATE: 2000-02-25
;  PRIOR APPLICATION NUMBER: 60/186,410
;  PRIOR FILING DATE: 2000-03-02
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
;   LENGTH: 488
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-794-822-11

  Query Match             25.0%;  Score 615;  DB 12;  Length 488;
  Best Local Similarity   35.0%;  Pred. No. 1.4e-45;
  Matches  179;  Conservative   74;  Mismatches  168;  Indels   90;  Gaps   18;

Qy         30 AFAAALGPLSFGFALGYSSPAIP------------SLQRAAPPAPRLDDAAASWFGAVVT 77
              |  ||||    ||  || :  |              |  ::     |   :    | ||:
Db          2 ALVAALGG---GFLFGYDTGVIGGFLALIDFLFRFGLLTSSGALAELVGYSTVLTGLVVS 58

Qy         78 ---LGAAAGGVLGGWLVDRAGRKLSLLLCSVPFVAGFAVITAAQ-----DVW---MLLGG 126
                 ||   | :  | | || ||| ||:   | || |  :  ||       :|   :|: |
Db         59 IFFLGRLIGSLFAGKLGDRFGRKKSLLIALVLFVIGALLSGAAPGYTTIGLWAFYLLIVG 118

Qy        127 RLLTGLACGVASLVAPVYISEIAYPAVRGLLGSCVQLMVVVGILLAYLAGWVL------- 179
              |:| ||  | ||:: |:||||||  |:|| |||  || : :|||:| : |  |
Db        119 RVLVGLGVGGASVLVPMYISEIAPKALRGALGSLYQLAITIGILVAAIIGLGLNKTNNDS 178

Qy        180 ---EWRWLAVLG--CVPPSLMLLLMCFMPETPRFLLTQHRRQEAMAALRFLWGSEQ-GWE 233
                  | |   ||   ||  |:|: : |:||:||:|: : : :||   |  | | |    |
```

```
Db        179 ALNSWGWRIPLGLQLVPALLLLIGLLFLPESPRWLVEKGKLEEAREVLAKLRGVEDVDQE 238

Qy        234 DPPIGAEQSFHLA-----------LLR---QPGIYKPFIIGVSLMAFQQLSGVNAVMFYA 279
                 | ||    ::           | |   :| : :  ::|| | |||||:|:||: :|:
Db        239 IQEIKAELEATVSEEKAGKASWGELFRGRTRPKVRQRLLMGVMLQAFQQLTGINAIFYYS 298

Qy        280 ETIFEEAKFKD---SSLASVVVGVIQVLFTAVAAL-IMDRAGRRLLLVLSGVVMVFSTSA 335
               |||:     |   | | :::|||:  :|| || : ::|| ||| ||:|    |
Db        299 PTIFKSVGVSDSVASLLVTIIVGVVNFVFTFVALIFLVDRFGRRPLLLLGAAGMAICFLI 358

Qy        336 FGAYFKLTQGGPGNSSHVAISAPVSAQPVDASVGLAWL-AVGSMCLFIAGFAVGWGPIPW 394
               ||              : ::  :  :| | |   | : |:  : |||| ||:|||||||
Db        359 LGA-------------SIGVALLLLNKPKDPSSKAAGIVAIVFILLFIAFFALGWGPIPW 405

Qy        395 LLMSEIFPLHVKGVATGICVLTNWLMAFLVTKEFSSLMEVLRPY--GAFWLA---SAFCI 449
              :::||:||  |:  |  :    |||  |::          | ||  ||  ||     | :
Db        406 VILSELFPTKVRSKALALATAANWLANFII--------GFLFPYITGAIGLALGGYVFLV 457

Qy        450 FS------VLFTLFCVPETKGKTLEQITAHF 474
              |:      :||  | ||||||:|||:|   |
Db        458 FAGLLVLFILFVFFFVPETKGRTLEEIEELF 488

RESULT 15
US-10-094-059-4
; Sequence 4, Application US/10094059
; Publication No. US20020127650A1
; GENERAL INFORMATION:
;  APPLICANT: Curtis, Rory A.J.
;  APPLICANT:  Millennium Pharmaceuticals Inc.
;  TITLE OF INVENTION: 32468, A Human Sugar Transporter Family Member and
;  TITLE OF INVENTION:  Uses Therefor
;  FILE REFERENCE: MPI01-040P1RM
;  CURRENT APPLICATION NUMBER: US/10/094,059
;  CURRENT FILING DATE:  2002-03-08
;  PRIOR APPLICATION NUMBER: 60/275,053
;  PRIOR FILING DATE: 2001-03-12
;  NUMBER OF SEQ ID NOS: 8
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
;   LENGTH: 488
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: consensus
US-10-094-059-4

  Query Match             25.0%;  Score 615;  DB 13;  Length 488;
  Best Local Similarity   35.0%;  Pred. No. 1.4e-45;
  Matches  179;  Conservative   74;  Mismatches  168;  Indels   90;  Gaps   18;

Qy         30 AFAAALGPLSFGFALGYSSPAIP------------SLQRAAPPAPRLDDAAASWFGAVVT 77
              |  ||||    ||  || :  |              |  ::     |   :    | ||:
Db          2 ALVAALGG---GFLFGYDTGVIGGFLALIDFLFRFGLLTSSGALAELVGYSTVLTGLVVS 58

Qy         78 ---LGAAAGGVLGGWLVDRAGRKLSLLLCSVPFVAGFAVITAAQ-----DVW---MLLGG 126
                 ||   | :  | | || ||| |||:  | || |  :  ||       :|   :|: |
Db         59 IFFLGRLIGSLFAGKLGDRFGRKKSLLIALVLFVIGALLSGAAPGYTTIGLWAFYLLIVG 118

Qy        127 RLLTGLACGVASLVAPVYISEIAYPAVRGLLGSCVQLMVVVGILLAYLAGWVL------- 179
              |:| ||  | ||:: |:||||||  |:|| |||  || : :|||:| : |  |
Db        119 RVLVGLGVGGASVLVPMYISEIAPKALRGALGSLYQLAITIGILVAAIIGLGLNKTNNDS 178

Qy        180 ---EWRWLAVLG--CVPPSLMLLLMCFMPETPRFLLTQHRRQEAMAALRFLWGSEQ-GWE 233
                  | |   ||   ||  |:|: : |:||:||:|: : : :||   |  | | |    |
Db        179 ALNSWGWRIPLGLQLVPALLLLIGLLFLPESPRWLVEKGKLEEAREVLAKLRGVEDVDQE 238

Qy        234 DPPIGAEQSFHLA-----------LLR---QPGIYKPFIIGVSLMAFQQLSGVNAVMFYA 279
                 | ||    ::           | |   :| : :  ::|| | |||||:|:||: :|:
Db        239 IQEIKAELEATVSEEKAGKASWGELFRGRTRPKVRQRLLMGVMLQAFQQLTGINAIFYYS 298

Qy        280 ETIFEEAKFKD---SSLASVVVGVIQVLFTAVAAL-IMDRAGRRLLLVLSGVVMVFSTSA 335
               |||:     |   | | :::|||:  :|| || : ::|| ||| ||:|    |
```

```
Db        299 PTIFKSVGVSDSVASLLVTIIVGVVNFVFTFVALIFLVDRFGRRPLLLLGAAGMAICFLI 358

Qy        336 FGAYFKLTQGGPGNSSHVAISAPVSAQPVDASVGLAWL-AVGSMCLFIAGFAVGWGPIPW 394
               ||              : ::  :  :| | |   | : |:  : |||| ||:|||||||
Db        359 LGA-------------SIGVALLLLNKPKDPSSKAAGIVAIVFILLFIAFFALGWGPIPW 405

Qy        395 LLMSEIFPLHVKGVATGICVLTNWLMAFLVTKEFSSLMEVLRPY--GAFWLA---SAFCI 449
              :::||:||  |:  |  :    |||  |::          | ||  ||  ||     | :
Db        406 VILSELFPTKVRSKALALATAANWLANFII--------GFLFPYITGAIGLALGGYVFLV 457

Qy        450 FS------VLFTLFCVPETKGKTLEQITAHF 474
              |:      :||  | ||||||:|||:|   |
Db        458 FAGLLVLFILFVFFFVPETKGRTLEEIEELF 488

Search completed: September 27, 2004, 19:41:42
Job time : 149 secs
```

```
                    GenCore version 5.1.6
           Copyright (c) 1993 - 2004  Compugen Ltd.


OM protein - protein search, using sw model

Run on:         September 27, 2004, 19:24:47 ; Search time 34 Seconds
                                           (without alignments)
                                           724.282 Million cell updates/sec

Title:          US-09-886-954A-1
Perfect score:  2457
Sequence:       1 MTPEDPEETQPLLGPPGGSA..........CVPETKGKTLEQITAHFEGR 477

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters:      389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Issued_Patents_AA:*
                1:  /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
                2:  /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
                3:  /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
                4:  /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
                5:  /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
                6:  /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                  SUMMARIES
                %
Result          Query
  No.   Score   Match Length DB  ID                         Description
 -------------------------------------------------------------------------
    1     615    25.0    488  4  US-10-162-012-46            Sequence 46, Appl
    2     605    24.6    514  4  US-09-489-039A-11902        Sequence 11902, A
    3   578.5    23.5    476  4  US-09-489-039A-11933        Sequence 11933, A
    4   575.5    23.4    494  2  US-09-031-392-5             Sequence 5, Appli
    5   575.5    23.4    494  3  US-09-299-549-5             Sequence 5, Appli
    6   575.5    23.4    494  4  US-09-610-417-5             Sequence 5, Appli
    7     568    23.1    493  2  US-09-031-392-10            Sequence 10, Appl
    8     568    23.1    493  3  US-09-299-549-10            Sequence 10, Appl
    9     568    23.1    493  4  US-09-610-417-10            Sequence 10, Appl
   10   548.5    22.3    518  4  US-09-679-686B-23           Sequence 23, Appl
   11   542.5    22.1    517  4  US-09-679-686B-18           Sequence 18, Appl
   12     536    21.8    509  2  US-09-031-392-6             Sequence 6, Appli
   13     536    21.8    509  3  US-09-299-549-6             Sequence 6, Appli
   14     536    21.8    509  4  US-09-610-417-6             Sequence 6, Appli
   15     535    21.8    514  4  US-09-679-686B-22           Sequence 22, Appl
   16   529.5    21.6    502  4  US-09-679-686B-2            Sequence 2, Appli
   17   526.5    21.4    519  4  US-09-679-686B-24           Sequence 24, Appl
   18   516.5    21.0    518  4  US-09-679-686B-16           Sequence 16, Appl
   19     516    21.0    584  2  US-08-928-692-13            Sequence 13, Appl
   20     516    21.0    584  4  US-09-339-972-13            Sequence 13, Appl
   21     513    20.9    524  2  US-08-928-692-12            Sequence 12, Appl
   22     513    20.9    524  4  US-09-339-972-12            Sequence 12, Appl
   23   507.5    20.7    501  4  US-09-489-039A-11731        Sequence 11731, A
   24   506.5    20.6    513  4  US-09-291-922-20            Sequence 20, Appl
   25   498.5    20.3    492  2  US-08-355-844-3             Sequence 3, Appli
   26   498.5    20.3    492  5  PCT-US95-16126-3            Sequence 3, Appli
   27     497    20.2    549  4  US-09-291-922-30            Sequence 30, Appl
   28     495    20.1    863  4  US-09-894-927B-9            Sequence 9, Appli
   29     489    19.9    521  4  US-09-489-039A-9549         Sequence 9549, Ap
   30   488.5    19.9    534  2  US-09-031-392-4             Sequence 4, Appli
   31   488.5    19.9    534  3  US-09-299-549-4             Sequence 4, Appli
   32   488.5    19.9    534  4  US-09-610-417-4             Sequence 4, Appli
   33   486.5    19.8    510  4  US-09-679-686B-19           Sequence 19, Appl
   34   485.5    19.8    488  2  US-08-928-692-10            Sequence 10, Appl
   35   485.5    19.8    488  4  US-09-339-972-10            Sequence 10, Appl
   36     478    19.5    562  4  US-10-162-012-44            Sequence 44, Appl
   37   465.5    18.9    539  4  US-09-291-922-26            Sequence 26, Appl
   38   464.5    18.9    523  4  US-09-291-922-24            Sequence 24, Appl
   39   458.5    18.7    529  4  US-09-291-922-28            Sequence 28, Appl
   40     454    18.5    383  2  US-09-031-392-3             Sequence 3, Appli
   41     454    18.5    383  3  US-09-299-549-3             Sequence 3, Appli
   42     454    18.5    383  4  US-09-610-417-3             Sequence 3, Appli
   43     444    18.1    510  4  US-09-291-922-22            Sequence 22, Appl
   44     437    17.8    511  4  US-09-679-686B-12           Sequence 12, Appl
   45   428.5    17.4    504  4  US-09-679-686B-21           Sequence 21, Appl


                                  ALIGNMENTS


RESULT 1
US-10-162-012-46
; Sequence 46, Application US/10162012
; Patent No. 6682597
; GENERAL INFORMATION:
;  APPLICANT: Curtis, Rory A.J.
;  APPLICANT:  Silos-Santiago, Inmaculada
;  APPLICANT:  Gu, Wei
;  TITLE OF INVENTION: NOVEL HUMAN ION CHANNEL AND TRANSPORTER FAMILY MEMBERS
;  FILE REFERENCE: 10448-190001
;  CURRENT APPLICATION NUMBER: US/10/162,012
;  CURRENT FILING DATE:  2002-06-04
;  PRIOR APPLICATION NUMBER: US 60/209,845
;  PRIOR FILING DATE: 2000-06-06
;  PRIOR APPLICATION NUMBER: US 09/875,321
;  PRIOR FILING DATE: 2001-06-06
;  PRIOR APPLICATION NUMBER: PCT/US01/18340
;  PRIOR FILING DATE: 2001-06-06
;  PRIOR APPLICATION NUMBER: US 60/209,257
;  PRIOR FILING DATE: 2000-06-05
;  PRIOR APPLICATION NUMBER: US 09/875,423
;  PRIOR FILING DATE: 2001-06-05
;  PRIOR APPLICATION NUMBER: PCT/US01/18398
;  PRIOR FILING DATE: 2001-06-05
;  PRIOR APPLICATION NUMBER: US 60/209,238
;  PRIOR FILING DATE: 2000-06-05
;  PRIOR APPLICATION NUMBER: US 09/875,363
;  PRIOR FILING DATE: 2001-06-05
;  PRIOR APPLICATION NUMBER: PCT/US01/18247
;  PRIOR FILING DATE: 2001-06-05
;  PRIOR APPLICATION NUMBER: US 60/227,068
;  PRIOR FILING DATE: 2000-08-22
;  PRIOR APPLICATION NUMBER: US 09/928,530
;  PRIOR FILING DATE: 2001-08-13
;  PRIOR APPLICATION NUMBER: PCT/US01/25475
;  PRIOR FILING DATE: 2001-08-15
;  PRIOR APPLICATION NUMBER: US 60/226,770
;  PRIOR FILING DATE: 2000-08-21
;  PRIOR APPLICATION NUMBER: US 09/934,421
;  PRIOR FILING DATE: 2001-08-21
;  PRIOR APPLICATION NUMBER: PCT/US01/26096
;  PRIOR FILING DATE: 2001-08-21
;  PRIOR APPLICATION NUMBER: US 60/279,281
;  PRIOR FILING DATE: 2001-03-28
;  PRIOR APPLICATION NUMBER: US 10/109,029
;  PRIOR FILING DATE: 2002-03-28
;  PRIOR APPLICATION NUMBER: PCT/US02/09728
;  PRIOR FILING DATE: 2002-03-28
;  PRIOR APPLICATION NUMBER: US 60/290,288
;  PRIOR FILING DATE: 2001-05-11
```

```
;  PRIOR APPLICATION NUMBER: US (not assigned)
;  PRIOR FILING DATE: 2002-05-13
;  NUMBER OF SEQ ID NOS: 48
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 46
;   LENGTH: 488
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: consensus sequence
US-10-162-012-46

  Query Match             25.0%;  Score 615;  DB 4;  Length 488;
  Best Local Similarity   35.0%;  Pred. No. 9.5e-50;
  Matches  179;  Conservative   74;  Mismatches  168;  Indels   90;  Gaps   18;

Qy           30 AFAAALGPLSFGFALGYSSPAIP------------SLQRAAPPAPRLDDAAASWFGAVVT 77
                |  ||||    ||  || :  |              |  ::     |   :    | ||:
Db            2 ALVAALGG---GFLFGYDTGVIGGFLALIDFLFRFGLLTSSGALAELVGYSTVLTGLVVS 58

Qy           78 ---LGAAAGGVLGGWLVDRAGRKLSLLLCSVPFVAGFAVITAAQ-----DVW---MLLGG 126
                   ||   | :  | | || ||| |||:  | || |  :  ||       :|   :|: |
Db           59 IFFLGRLIGSLFAGKLGDRFGRKKSLLIALVLFVIGALLSGAAPGYTTIGLWAFYLLIVG 118

Qy          127 RLLTGLACGVASLVAPVYISEIAYPAVRGLLGSCVQLMVVVGILLAYLAGWVL------- 179
                |:| ||  | ||:: |:||||||  |:|| |||  || : :|||:| : |  |
Db          119 RVLVGLGVGGASVLVPMYISEIAPKALRGALGSLYQLAITIGILVAAIIGLGLNKTNNDS 178

Qy          180 ---EWRWLAVLG--CVPPSLMLLLMCFMPETPRFLLTQHRRQEAMAALRFLWGSEQ-GWE 233
                    | |   ||   ||  |:|: : |:||:||:|: : : :||   |  | | |    |
Db          179 ALNSWGWRIPLGLQLVPALLLLIGLLFLPESPRWLVEKGKLEEAREVLAKLRGVEDVDQE 238

Qy          234 DPPIGAEQSFHLA-----------LLR---QPGIYKPFIIGVSLMAFQQLSGVNAVMFYA 279
                   | ||   ::            | |   :| : :  ::|| | |||||:|:||: :|:
Db          239 IQEIKAELEATVSEEKAGKASWGELFRGRTRPKVRQRLLMGVMLQAFQQLTGINAIFYYS 298

Qy          280 ETIFEEAKFKD---SSLASVVVGVIQVLFTAVAAL-IMDRAGRRLLLVLSGVVMVFSTSA 335
                 |||:     |   | | :::|||:  :|| || : ::|| ||| ||:|    |
Db          299 PTIFKSVGVSDSVASLLVTIIVGVVNFVFTFVALIFLVDRFGRRPLLLLGAAGMAICFLI 358

Qy          336 FGAYFKLTQGGPGNSSHVAISAPVSAQPVDASVGLAWL-AVGSMCLFIAGFAVGWGPIPW 394
                 ||              : ::  :  :| | |   | : |:  : |||| ||:|||||||
Db          359 LGA-------------SIGVALLLLNKPKDPSSKAAGIVAIVFILLFIAFFALGWGPIPW 405

Qy          395 LLMSEIFPLHVKGVATGICVLTNWLMAFLVTKEFSSLMEVLRPY--GAFWLA---SAFCI 449
                :::||:||  |:  |  :    |||  |::          | ||  ||  ||     | :
Db          406 VILSELFPTKVRSKALALATAANWLANFII--------GFLFPYITGAIGLALGGYVFLV 457

Qy          450 FS------VLFTLFCVPETKGKTLEQITAHF 474
                |:      :||  | ||||||:|||:|   |
Db          458 FAGLLVLFILFVFFFVPETKGRTLEEIEELF 488

RESULT 2
US-09-489-039A-11902
; Sequence 11902, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
;  APPLICANT:  Gary Breton et. al
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
;  TITLE OF INVENTION:  PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  2709.2004001
;  CURRENT APPLICATION NUMBER: US/09/489,039A
;  CURRENT FILING DATE:  2000-01-27
;  PRIOR APPLICATION NUMBER:  US 60/117,747
;  PRIOR FILING DATE:  1999-01-29
;  NUMBER OF SEQ ID NOS:  14342
; SEQ ID NO 11902
;   LENGTH: 514
;   TYPE: PRT
;   ORGANISM: Klebsiella pneumoniae
US-09-489-039A-11902

  Query Match             24.6%;  Score 605;  DB 4;  Length 514;
  Best Local Similarity   33.3%;  Pred. No. 9.1e-49;
  Matches  156;  Conservative   86;  Mismatches  187;  Indels   40;  Gaps    9;

Qy           20 APRGRR-----VFLAAFAAALGPLSFGFALGYSSPAIPSLQRAAPPAPRLDDAAASWFGA 74
                :|| :|      :  : |||:  | ||  :|  | |:| :      : :| :    |  :
Db           52 SPRTQRDTRRMNWFVSIAAAVAGLLFGLDIGVISGALPFITDHFTLSSQLQE----WVVS 107

Qy           75 VVTLGAAAGGVLGGWLVDRAGRKLSLLLCSVPFVAGFAVITAAQDVWMLLGGRLLTGLAC 134
                 : |||| | :  |||  | ||| ||:  :| ||||      |  | :||  |:: |:|
Db          108 SMMLGAAIGALFNGWLSFRLGRKYSLMAGAVLFVAGSIGSAFAASVEVLLVARVVLGVAV 167

Qy          135 GVASLVAPVYISEIAYPAVRGLLGSCVQLMVVVGILLAYLAGWVLEW--RWLAVLG--CV 190
                |:||  ||:|:||:|   ||| : |  |||| :||:||:|:     :   | |:||   :
Db          168 GIASYTAPLYLSEMASENVRGKMISMYQLMVTLGIVLAFLSDTAFSYSGNWRAMLGVLAL 227

Qy          191 PPSLMLLLMCFMPETPRFLLTQHRRQEAMAALRFLWGSEQGWEDPPIGAEQSFHL----- 245
                |  ::::|: |:| :||:|  : |  ||   || |  : :   |      :|  |
Db          228 PAVILIILVVFLPNSPRWLAEKGRHIEAEEVLRMLRDTSEKARDELNEIRESLKLKQGGW 287

Qy          246 ALLR-QPGIYKPFIIGVSLMAFQQLSGVNAVMFYAETIFEEAKF---KDSSLASVVVGVI 301
                || :    : :   :|: | | || :|:| :|:||  ||: | |   :   :|::|||:
Db          288 ALFKINRNVRRAVFLGMLLQAMQQFTGMNIIMYYAPRIFKMAGFTTTEQQMIATLVVGLT 347

Qy          302 QVLFTAVAALIMDRAGRRLLLVLSGVVMVFSTSAFGAYFKLTQGGPGNSSHVAISAPVSA 361
                 :  | :|   :|:|||:  | :   ||   |   |  : | |   |
Db          348 FMFATFIAVFTVDKAGRKPALKIGFSVMALGTLVLG--YCLMQFDNGT------------ 393

Qy          362 QPVDASVGLAWLAVGSMCLFIAGFAVGWGPIPWLLMSEIFPLHVKGVATGICVLTNWLMA 421
                    || ||:||:||   : |||:|:   |: |:| ||| ||  :         |||:
Db          394 ----ASSGLSWLSVGMTMMCIAGYAMSAAPVVWILCSEIQPLKCRDFGITCSTTTNWVSN 449

Qy          422 FLVTKEFSSLMEVLRPYGAFWLASAFCIFSVLFTLFCVPETKGKTLEQI 470
                 ::   | :|:: :   | ||| :|  :  :  | : :||||  ||| |
Db          450 MIIGATFLTLLDAIGAAGTFWLYTALNVAFIGITFWLIPETKNVTLEHI 498

RESULT 3
US-09-489-039A-11933
; Sequence 11933, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
;  APPLICANT:  Gary Breton et. al
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
;  TITLE OF INVENTION:  PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  2709.2004001
;  CURRENT APPLICATION NUMBER: US/09/489,039A
;  CURRENT FILING DATE:  2000-01-27
;  PRIOR APPLICATION NUMBER:  US 60/117,747
;  PRIOR FILING DATE:  1999-01-29
;  NUMBER OF SEQ ID NOS:  14342
; SEQ ID NO 11933
;   LENGTH: 476
;   TYPE: PRT
;   ORGANISM: Klebsiella pneumoniae
US-09-489-039A-11933

  Query Match             23.5%;  Score 578.5;  DB 4;  Length 476;
  Best Local Similarity   31.2%;  Pred. No. 2.6e-46;
  Matches  149;  Conservative   84;  Mismatches  180;  Indels   65;  Gaps   10;

Qy           27 FLAAFAAALGPLSFGFALGYSSPAIPSLQRAAPPAPRLDDAAASWFGAVVTLGAAAGGVL 86
                |   | |||  | ||  :|  : |:| :        ::      |  : :  ||| | |
Db           28 FFVCFLAALAGLLFGLDIGVIAGALPFIANEF----QISAHTQEWVVSSMMFGAAVGAVG 83

Qy           87 GGWLVDRAGRKLSLLLCSVPFVAGFAVITAAQDVWMLLGGRLLTGLACGVASLVAPVYIS 146
                 |||  : ||| ||:: :: ||||     || :| :||  |:| ||| ||||  ||:|:|
Db           84 SGWLSFKLGRKKSLMIGAILFVAGSLFSAAAPNVEILLVSRVLLGLAVGVASYTAPLYLS 143

Qy          147 EIAYPAVRGLLGSCVQLMVVVGILLAYLAGWVLE----WRWLAVLGCVPPSLMLLLMCFM 202
                |||   :|| : |  |||: :||| |||:         |||:  :  :|  |:|: : |:
Db          144 EIAPEKIRGSMISMYQLMITIGILGAYLSDTAFSYSGAWRWMLGVIIIPAVLLLIGVIFL 203
```

```
Qy        203 PETPRFLLTQHRRQEAMAAL------------------RFLWGSEQGWEDPPIGAEQSFH 244
              |::||:   : |  :|   |                    |   : ||
Db        204 PDSPRWFAAKRRFVDAERVLLRLRDTSAEAKRELDEIRESLKVKQSGW------------ 251

Qy        245 LALLRQPGIYKPFI-IGVSLMAFQQLSGVNAVMFYAETIFEEAKFKDSS---LASVVVGV 300
               :| :    ::  : :|: |   || :|:| :|:||  ||| | : :::     :|:||:
Db        252 -SLFKDNSNFRRAVFLGILLQVMQQFTGMNVIMYYAPKIFELAGYANTTEQMWGTVIVGL 310

Qy        301 IQVLFTAVAALIMDRAGRRLLLVLSGVVMVFSTSAFGAYFKLTQGGPGNSSHVAISAPVS 360
                || | :|  ::|| ||:  |:|  :||     | |       |  |   |: | :  :
Db        311 TNVLATFIAIGLVDRWGRKPTLILGFIVM-----AAG------MGVLGTMMHIGIHSSTA 359

Qy        361 AQPVDASVGLAWLAVGSMCLFIAGFAVGWGPIPWLLMSEIFPLHVKGVATGICVLTNWLM 420
                         ::||  : :|| |||:  ||: |:| ||| ||  :         |||:
Db        360 ----------QYIAVLMLLMFIVGFAMSAGPLIWVLCSEIQPLKGRDFGITCSTATNWIA 409

Qy        421 AFLVTKEFSSLMEVLRPYGAFWLASAFCIFSVLFTLFCVPETKGKTLEQITAH-FEGR 477
                :|   | :::  |     ||:     :  :| ||: :||||  :|| |  :  :||
Db        410 NMIVGATFLTMLNSLGSANTFWVYGGLNVLFILLTLWLIPETKNVSLEHIERNLMQGR 467

RESULT 4
US-09-031-392-5
; Sequence 5, Application US/09031392
; Patent No. 5942398
;  GENERAL INFORMATION:
;    APPLICANT:  Tartaglia, Louis A.
;    APPLICANT:  Weng, Xun
;    TITLE OF INVENTION:  NUCLEIC ACID MOLECULES
;    TITLE OF INVENTION:  ENCODING GLUTEX AND USES THEREOF
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Windows95
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/031,392
;      FILING DATE:  26-FEB-1998
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meiklejohn, Ph.D., Anita L.
;      REGISTRATION NUMBER:  35,283
;      REFERENCE/DOCKET NUMBER:  07334/072001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/542-5070
;      TELEFAX:  617/542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  494 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-031-392-5

  Query Match             23.4%;  Score 575.5;  DB 2;  Length 494;
  Best Local Similarity   30.7%;  Pred. No. 5.3e-46;
  Matches  150;  Conservative   94;  Mismatches  179;  Indels   65;  Gaps   10;

Qy         26 VFLAAFAAALGPLSFGFALGYSSPAIPSL----------QRAAPPAPRLDDAAASWFGAV 75
              :  |   | :|   ||:  |  :     :          :   ||:  |  :  |   |:
Db         10 LIFAISIATIGSFQFGYNTGVINAPEAIIKDFLNYTLEERSETPPSSVLLTSLWSLSVAI 69

Qy         76 VTLGAAAGGVLGGWLVDRAGRKLSLLLCSVPFVAGFAVI---TAAQDVWMLLGGRLLTGL 132
               ::|   |    |  |:| ||: |:|: ::  :||  ::     |: | ||: |||: ||
Db         70 FSVGGMIGSFSVGLFVNRFGRRNSMLIVNLLAIAGGCLMGFCKIAESVEMLILGRLIIGL 129
```

```
Qy        133 ACGVASLVAPVYISEIAYPAVRGLLGSCVQLMVVVGILLAYLAGWVL----EWRWLAVLG 188
               ||: :   |:|| ||:  |:||  |:  || :|:|||:| : |  :    |  |  :||
Db        130 FCGLCTGFVPMYIGEISPTALRGAFGTLNQLGIVIGILVAQIFGLKVILGTEDLWPLLLG 189

Qy        189 --CVPPSLMLLLMCFMPETPRFLLTQHRRQE-AMAALRFLWGSEQGWEDPPIGAEQSFH- 244
                 :|  :    : | ||:|||||   : :| |   |: |||:|   :|     ::|
Db        190 FTILPAIIQCAALPFCPESPRFLLINRKEEEKAKEILQRLWGTEDVAQDIQEMKDESMRM 249

Qy        245 --------LALLRQPGIYKPFIIGVSLMAFQQLSGVNAVMFYAETIFEEAKFKDSSLASV 296
                      | | | |   :| || : |   |||||:||| :|:  ||::|  ::   |::
Db        250 SQEKQVTVLELFRAPNYRQPIIISIMLQLSQQLSGINAVFYYSTGIFKDAGVQEPVYATI 309

Qy        297 VVGVIQVLFTAVAALIMDRAGRRLLLVLSGVVMVFSTSAFGAYFKLTQGGPGNSSHVAIS 356
                ||:  :|| |:  :::||||| | ::                     | |  :  :|
Db        310 GAGVVNTIFTVVSVFLVERAGRRTLHLI---------------------GLGGMAFCSIL 348

Qy        357 APVSAQPVDASVGLAWLAVGSMCLFIAGFAVGWGPIPWLLMSEIFPLHVKGVATGICVLT 416
                :|    |    :::: :|:: :|:| | :| ||||| :::|:|    :  |  :   :
Db        349 MTISLLLKDNYSWMSFICIGAILVFVAFFEIGPGPIPWFIVAELFGQGPRPAAMAVAGCS 408

Qy        417 NWLMAFLVTKEFSSLMEVLRPYGAFWL-ASAFCIFSV------LFTLFCVPETKGKTLEQ 469
              ||   |||         :| |   |:| |  | :|:|      :|| | ||||:|:| |:
Db        409 NWTSNFLV--------GLLFPSATFYLGAYVFIVFTVFLVIFWVFTFFKVPETRGRTFEE 460

Qy        470 ITAHFEGR 477
              ||  |||:
Db        461 ITRAFEGQ 468

RESULT 5
US-09-299-549-5
; Sequence 5, Application US/09299549
; Patent No. 6136547
;  GENERAL INFORMATION:
;    APPLICANT:  Tartaglia, Louis A.
;    APPLICANT:  Weng, Xun
;    TITLE OF INVENTION:  NUCLEIC ACID MOLECULES ENCODING
;    TITLE OF INVENTION:  GLUTEX AND USES THEREOF
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Windows95
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/299,549
;      FILING DATE:  26-APR-1999
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  09/031,392
;      FILING DATE:  26-FEB-1998
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meiklejohn, Ph.D., Anita L.
;      REGISTRATION NUMBER:  35,283
;      REFERENCE/DOCKET NUMBER:  07334/072002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/542-5070
;      TELEFAX:  617/542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  494 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-299-549-5
```

```
  Query Match             23.4%;  Score 575.5;  DB 3;  Length 494;
  Best Local Similarity   30.7%;  Pred. No. 5.3e-46;
  Matches  150;  Conservative   94;  Mismatches  179;  Indels   65;  Gaps   10;

Qy          26 VFLAAFAAALGPLSFGFALGYSSPAIPSL----------QRAAPPAPRLDDAAASWFGAV 75
               :  |   | :|   ||:  |  :     :          :   ||:  |  :  |   |:
Db          10 LIFAISIATIGSFQFGYNTGVINAPEAIIKDFLNYTLEERSETPPSSVLLTSLWSLSVAI 69

Qy          76 VTLGAAAGGVLGGWLVDRAGRKLSLLLCSVPFVAGFAVI---TAAQDVWMLLGGRLLTGL 132
                ::|   |    |  |:| ||: |:|: ::  :||  ::     |: | ||: |||: ||
Db          70 FSVGGMIGSFSVGLFVNRFGRRNSMLIVNLLAIAGGCLMGFCKIAESVEMLILGRLIIGL 129

Qy         133 ACGVASLVAPVYISEIAYPAVRGLLGSCVQLMVVVGILLAYLAGWVL----EWRWLAVLG 188
                ||: :   |:|| ||:  |:||  |:  || :|:||||:| : |  :    |  |  :||
Db         130 FCGLCTGFVPMYIGEISPTALRGAFGTLNQLGIVIGILVAQIFGLKVILGTEDLWPLLLG 189

Qy         189 --CVPPSLMLLLMCFMPETPRFLLTQHRRQE-AMAALRFLWGSEQGWEDPPIGAEQSFH- 244
                  :|  :    : | ||:|||||   : :| |   |: |||:|   :|     ::|
Db         190 FTILPAIIQCAALPFCPESPRFLLINRKEEEKAKEILQRLWGTEDVAQDIQEMKDESMRM 249

Qy         245 --------LALLRQPGIYKPFIIGVSLMAFQQLSGVNAVMFYAETIFEEAKFKDSSLASV 296
                       | | | |   :| || : |   |||||:||| :|:  ||::|  ::   |::
Db         250 SQEKQVTVLELFRAPNYRQPIIISIMLQLSQQLSGINAVFYYSTGIFKDAGVQEPVYATI 309

Qy         297 VVGVIQVLFTAVAALIMDRAGRRLLLVLSGVVMVFSTSAFGAYFKLTQGGPGNSSHVAIS 356
                 ||:  :|| |:  :::||||| | ::                     | |  :  :|
Db         310 GAGVVNTIFTVVSVFLVERAGRRTLHLI---------------------GLGGMAFCSIL 348

Qy         357 APVSAQPVDASVGLAWLAVGSMCLFIAGFAVGWGPIPWLLMSEIFPLHVKGVATGICVLT 416
                 :|    |    :::: :|:: :|:| | :| ||||| :::|:|    :  |  :   :
Db         349 MTISLLLKDNYSWMSFICIGAILVFVAFFEIGPGPIPWFIVAELFGQGPRPAAMAVAGCS 408

Qy         417 NWLMAFLVTKEFSSLMEVLRPYGAFWL-ASAFCIFSV------LFTLFCVPETKGKTLEQ 469
               ||   |||         :| |   |:| |  | :|:|      :|| | ||||:|:| |:
Db         409 NWTSNFLV--------GLLFPSATFYLGAYVFIVFTVFLVIFWVFTFFKVPETRGRTFEE 460

Qy         470 ITAHFEGR 477
               ||  |||:
Db         461 ITRAFEGQ 468

RESULT 6
US-09-610-417-5
; Sequence 5, Application US/09610417
; Patent No. 6346374
;  GENERAL INFORMATION:
;       APPLICANT: Tartaglia, Louis A.
;                  Weng, Xun
;       TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
;                           GLUTEX AND USES THEREOF
;       NUMBER OF SEQUENCES: 10
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Fish & Richardson P.C.
;            STREET: 225 Franklin Street
;            CITY: Boston
;            STATE: MA
;            COUNTRY: USA
;            ZIP: 02110-2804
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Diskette
;            COMPUTER: IBM Compatible
;            OPERATING SYSTEM: Windows95
;            SOFTWARE: FastSEQ for Windows Version 2.0
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/610,417
;            FILING DATE: 05-Jul-2000
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 09/299,549
;            FILING DATE: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Meiklejohn, Ph.D., Anita L.
;            REGISTRATION NUMBER: 35,283
```

```
;            REFERENCE/DOCKET NUMBER: 07334/072002
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 617/542-5070
;            TELEFAX: 617/542-8906
;            TELEX: 200154
;  INFORMATION FOR SEQ ID NO: 5:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 494 amino acids
;            TYPE: amino acid
;            TOPOLOGY: linear
;       MOLECULE TYPE: protein
;       SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-610-417-5

  Query Match             23.4%;  Score 575.5;  DB 4;  Length 494;
  Best Local Similarity   30.7%;  Pred. No. 5.3e-46;
  Matches  150;  Conservative   94;  Mismatches  179;  Indels   65;  Gaps   10;

Qy          26 VFLAAFAAALGPLSFGFALGYSSPAIPSL----------QRAAPPAPRLDDAAASWFGAV 75
               :  |   | :|   ||:  |  :     :          :   ||:  |  :  |   |:
Db          10 LIFAISIATIGSFQFGYNTGVINAPEAIIKDFLNYTLEERSETPPSSVLLTSLWSLSVAI 69

Qy          76 VTLGAAAGGVLGGWLVDRAGRKLSLLLCSVPFVAGFAVI---TAAQDVWMLLGGRLLTGL 132
                ::|   |    |  |:| ||: |:|: ::  :||  ::     |: | ||: |||: ||
Db          70 FSVGGMIGSFSVGLFVNRFGRRNSMLIVNLLAIAGGCLMGFCKIAESVEMLILGRLIIGL 129

Qy         133 ACGVASLVAPVYISEIAYPAVRGLLGSCVQLMVVVGILLAYLAGWVL----EWRWLAVLG 188
                ||: :   |:|| ||:  |:||  |:  || :|:||||:| : |  :    |  |  :||
Db         130 FCGLCTGFVPMYIGEISPTALRGAFGTLNQLGIVIGILVAQIFGLKVILGTEDLWPLLLG 189

Qy         189 --CVPPSLMLLLMCFMPETPRFLLTQHRRQE-AMAALRFLWGSEQGWEDPPIGAEQSFH- 244
                  :|  :    : | ||:|||||   : :| |   |: |||:|   :|     ::|
Db         190 FTILPAIIQCAALPFCPESPRFLLINRKEEEKAKEILQRLWGTEDVAQDIQEMKDESMRM 249

Qy         245 --------LALLRQPGIYKPFIIGVSLMAFQQLSGVNAVMFYAETIFEEAKFKDSSLASV 296
                       | | | |   :| || : |   |||||:||| :|:  ||::|  ::   |::
Db         250 SQEKQVTVLELFRAPNYRQPIIISIMLQLSQQLSGINAVFYYSTGIFKDAGVQEPVYATI 309

Qy         297 VVGVIQVLFTAVAALIMDRAGRRLLLVLSGVVMVFSTSAFGAYFKLTQGGPGNSSHVAIS 356
                 ||:  :|| |:  :::||||| | ::                     | |  :  :|
Db         310 GAGVVNTIFTVVSVFLVERAGRRTLHLI---------------------GLGGMAFCSIL 348

Qy         357 APVSAQPVDASVGLAWLAVGSMCLFIAGFAVGWGPIPWLLMSEIFPLHVKGVATGICVLT 416
                 :|    |    :::: :|:: :|:| | :| ||||| :::|:|    :  |  :   :
Db         349 MTISLLLKDNYSWMSFICIGAILVFVAFFEIGPGPIPWFIVAELFGQGPRPAAMAVAGCS 408

Qy         417 NWLMAFLVTKEFSSLMEVLRPYGAFWL-ASAFCIFSV------LFTLFCVPETKGKTLEQ 469
               ||   |||         :| |   |:| |  | :|:|      :|| | ||||:|:| |:
Db         409 NWTSNFLV--------GLLFPSATFYLGAYVFIVFTVFLVIFWVFTFFKVPETRGRTFEE 460

Qy         470 ITAHFEGR 477
               ||  |||:
Db         461 ITRAFEGQ 468

RESULT 7
US-09-031-392-10
; Sequence 10, Application US/09031392
; Patent No. 5942398
;  GENERAL INFORMATION:
;    APPLICANT:  Tartaglia, Louis A.
;    APPLICANT:  Weng, Xun
;    TITLE OF INVENTION:  NUCLEIC ACID MOLECULES
;    TITLE OF INVENTION:  ENCODING GLUTEX AND USES THEREOF
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
```

```
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Windows95
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/031,392
;      FILING DATE:  26-FEB-1998
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meiklejohn, Ph.D., Anita L.
;      REGISTRATION NUMBER:  35,283
;      REFERENCE/DOCKET NUMBER:  07334/072001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/542-5070
;      TELEFAX:  617/542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  493 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-031-392-10

  Query Match             23.1%;  Score 568;  DB 2;  Length 493;
  Best Local Similarity   32.2%;  Pred. No. 2.7e-45;
  Matches  159;  Conservative   87;  Mismatches  160;  Indels   88;  Gaps   14;

Qy         28 LAAFAAALGPLSFGFALGYSSPAIPSLQRAAPPAPRLDDAAASWFG------AVVTLGA- 80
              ||   ||||   :|: ||  :           |   ::    :| |      :| ||
Db         16 LAVLIAALGSFQYGYNLGVIN----------APQKVIEAFYETWLGRXGEXPSVPTLTLL 65

Qy         81 --------AAGGVLGGWLV----DRAGRKLSLLLCSVPFVAG---FAVITAAQDVWMLLG 125
                      | ||::| :||    :| ||| ::|: :|  :||     :   |    ||:
Db         66 WSLSVSIFAVGGMIGSFLVGXIGNRLGRKXAMLVNNVLAIAGGLLMGLAKXAXSFEMLIL 125

Qy        126 GRLLTGLACGVASLVAPVYISEIAYPAVRGLLGSCVQLMVVVGILLAYLAGWVLE----- 180
              || : || ||::| | |:|: ||:  |:|| ||:  || :|:||:|  : |  |:
Db        126 GRFIIGLYCGLSSGVVPMYVGEISPTALRGALGTLNQLGIVIGILIAQVLG--LDSLLGN 183

Qy        181 ---WRWLAVLGCVPPSLMLLLMCFMPETPRFLLTQHRRQEAMA--ALRFLWG-------- 227
                 |  |  |  ||  | |||: | ||:||:||  :: :|| |  ||: | |
Db        184 ESLWPLLLGLTGVPALLQLLLLPFCPESPRYLLI-NKNEEARAKKALQRLRGTADVSQEV 242

Qy        228 SEQGWEDPPIGAEQSFH-LALLRQPGIYKPFIIGVSLMAFQQLSGVNAVMFYAETIFEEA 286
              :|   |   : :|:    | | |     :| || : |   |||||:||| :|: :|||:|
Db        243 AEMKDESRXMXSEKXVSVLELFRSRXYRQPVIIAIVLQLSQQLSGINAVFYYSTSIFEKA 302

Qy        287 KFKDSSLASVVVGVIQVLFTAVAALIMDRAGRRLLLVLSGVVMVFSTSAFGAYFKLTQGG 346
                     |::  ||:  :|| |:  :::||||| | :|                     |
Db        303 GVGQPVYATIGAGVVNTVFTVVSVFVVERAGRRTLHLL---------------------G 341

Qy        347 PGNSSHVAISAPVSAQPVDASVGLAWLAVGSMCLFIAGFAVGWGPIPWLLMSEIFPLHVK 406
               |  :  |:   ::   :|    :::::: ::  |:| | || ||||| :::|:|    :
Db        342 LGGMAGCAVLMTIALALLDQVPWMSYVSIVAIFGFVAFFEVGPGPIPWFIVAELFSQGPR 401

Qy        407 GVATGICVLTNWLMAFLVTKEFSSLMEVLRPYGAFWLASAFCIFSVL------FTLFCVP 460
                |  :   :||   |:|   |  : |:| ||        | :|:||      ||   ||
Db        402 PAAIAVAGFSNWTSNFIVGLLFQYIAELLGPY-------VFIVFAVLLLLFFIFTFLKVP 454

Qy        461 ETKGKTLEQITAHF 474
              ||||:| ::| | |
Db        455 ETKGRTFDEIAAAF 468

RESULT 8
US-09-299-549-10
; Sequence 10, Application US/09299549
; Patent No. 6136547
;  GENERAL INFORMATION:
;    APPLICANT:  Tartaglia, Louis A.
;    APPLICANT:  Weng, Xun
;    TITLE OF INVENTION:  NUCLEIC ACID MOLECULES ENCODING
;    TITLE OF INVENTION:  GLUTEX AND USES THEREOF
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Windows95
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/299,549
;      FILING DATE:  26-APR-1999
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  09/031,392
;      FILING DATE:  26-FEB-1998
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meiklejohn, Ph.D., Anita L.
;      REGISTRATION NUMBER:  35,283
;      REFERENCE/DOCKET NUMBER:  07334/072002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/542-5070
;      TELEFAX:  617/542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  493 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-299-549-10

  Query Match             23.1%;  Score 568;  DB 3;  Length 493;
  Best Local Similarity   32.2%;  Pred. No. 2.7e-45;
  Matches  159;  Conservative   87;  Mismatches  160;  Indels   88;  Gaps   14;

Qy         28 LAAFAAALGPLSFGFALGYSSPAIPSLQRAAPPAPRLDDAAASWFG------AVVTLGA- 80
              ||   ||||   :|: ||  :           |   ::    :| |      :| ||
Db         16 LAVLIAALGSFQYGYNLGVIN----------APQKVIEAFYETWLGRXGEXPSVPTLTLL 65

Qy         81 --------AAGGVLGGWLV----DRAGRKLSLLLCSVPFVAG---FAVITAAQDVWMLLG 125
                      | ||::| :||    :| ||| ::|: :|  :||     :   |    ||:
Db         66 WSLSVSIFAVGGMIGSFLVGXIGNRLGRKXAMLVNNVLAIAGGLLMGLAKXAXSFEMLIL 125

Qy        126 GRLLTGLACGVASLVAPVYISEIAYPAVRGLLGSCVQLMVVVGILLAYLAGWVLE----- 180
              || : || ||::| | |:|: ||:  |:|| ||:  || :|:||:|  : |  |:
Db        126 GRFIIGLYCGLSSGVVPMYVGEISPTALRGALGTLNQLGIVIGILIAQVLG--LDSLLGN 183

Qy        181 ---WRWLAVLGCVPPSLMLLLMCFMPETPRFLLTQHRRQEAMA--ALRFLWG-------- 227
                 |  |  |  ||  | |||: | ||:||:||  :: :|| |  ||: | |
Db        184 ESLWPLLLGLTGVPALLQLLLLPFCPESPRYLLI-NKNEEARAKKALQRLRGTADVSQEV 242

Qy        228 SEQGWEDPPIGAEQSFH-LALLRQPGIYKPFIIGVSLMAFQQLSGVNAVMFYAETIFEEA 286
              :|   |   : :|:    | | |     :| || : |   |||||:||| :|: :|||:|
Db        243 AEMKDESRXMXSEKXVSVLELFRSRXYRQPVIIAIVLQLSQQLSGINAVFYYSTSIFEKA 302

Qy        287 KFKDSSLASVVVGVIQVLFTAVAALIMDRAGRRLLLVLSGVVMVFSTSAFGAYFKLTQGG 346
                     |::  ||:  :|| |:  :::||||| | :|                     |
Db        303 GVGQPVYATIGAGVVNTVFTVVSVFVVERAGRRTLHLL---------------------G 341

Qy        347 PGNSSHVAISAPVSAQPVDASVGLAWLAVGSMCLFIAGFAVGWGPIPWLLMSEIFPLHVK 406
               |  :  |:   ::   :|    :::::: ::  |:| | || ||||| :::|:|    :
Db        342 LGGMAGCAVLMTIALALLDQVPWMSYVSIVAIFGFVAFFEVGPGPIPWFIVAELFSQGPR 401

Qy        407 GVATGICVLTNWLMAFLVTKEFSSLMEVLRPYGAFWLASAFCIFSVL------FTLFCVP 460
                |  :   :||   |:|   |  : |:| ||        | :|:||      ||   ||
Db        402 PAAIAVAGFSNWTSNFIVGLLFQYIAELLGPY-------VFIVFAVLLLLFFIFTFLKVP 454

Qy        461 ETKGKTLEQITAHF 474
              ||||:| ::| | |
```

```
Db        455 ETKGRTFDEIAAAF 468

RESULT 9
US-09-610-417-10
; Sequence 10, Application US/09610417
; Patent No. 6346374
;  GENERAL INFORMATION:
;       APPLICANT: Tartaglia, Louis A.
;                  Weng, Xun
;       TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
;                           GLUTEX AND USES THEREOF
;       NUMBER OF SEQUENCES: 10
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Fish & Richardson P.C.
;            STREET: 225 Franklin Street
;            CITY: Boston
;            STATE: MA
;            COUNTRY: USA
;            ZIP: 02110-2804
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Diskette
;            COMPUTER: IBM Compatible
;            OPERATING SYSTEM: Windows95
;            SOFTWARE: FastSEQ for Windows Version 2.0
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/610,417
;            FILING DATE: 05-Jul-2000
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 09/299,549
;            FILING DATE: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Meiklejohn, Ph.D., Anita L.
;            REGISTRATION NUMBER: 35,283
;            REFERENCE/DOCKET NUMBER: 07334/072002
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 617/542-5070
;            TELEFAX: 617/542-8906
;            TELEX: 200154
;  INFORMATION FOR SEQ ID NO: 10:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 493 amino acids
;            TYPE: amino acid
;            TOPOLOGY: linear
;       MOLECULE TYPE: protein
;       SEQUENCE DESCRIPTION: SEQ ID NO: 10:
US-09-610-417-10

  Query Match             23.1%;  Score 568;  DB 4;  Length 493;
  Best Local Similarity   32.2%;  Pred. No. 2.7e-45;
  Matches  159;  Conservative   87;  Mismatches  160;  Indels   88;  Gaps   14;

Qy         28 LAAFAAALGPLSFGFALGYSSPAIPSLQRAAPPAPRLDDAAASWFG------AVVTLGA- 80
              ||   ||||   :|: ||  :           |   ::    :| |      :| ||
Db         16 LAVLIAALGSFQYGYNLGVIN----------APQKVIEAFYETWLGRXGEXPSVPTLTLL 65

Qy         81 --------AAGGVLGGWLV----DRAGRKLSLLLCSVPFVAG---FAVITAAQDVWMLLG 125
                      | ||::| :||    :| ||| ::|: :|  :||     :   |    ||:
Db         66 WSLSVSIFAVGGMIGSFLVGXIGNRLGRKXAMLVNNVLAIAGGLLMGLAKXAXSFEMLIL 125

Qy        126 GRLLTGLACGVASLVAPVYISEIAYPAVRGLLGSCVQLMVVVGILLAYLAGWVLE----- 180
              || : || ||::| | |:|: ||:  |:|| ||:  || :|:|||:| : |  |:
Db        126 GRFIIGLYCGLSSGVVPMYVGEISPTALRGALGTLNQLGIVIGILIAQVLG--LDSLLGN 183

Qy        181 ---WRWLAVLGCVPPSLMLLLMCFMPETPRFLLTQHRRQEAMA--ALRFLWG-------- 227
                 |  |  |  ||  | |||: | ||:||:||  :: :|| |  ||: | |
Db        184 ESLWPLLLGLTGVPALLQLLLLPFCPESPRYLLI-NKNEEARAKKALQRLRGTADVSQEV 242

Qy        228 SEQGWEDPPIGAEQSFH-LALLRQPGIYKPFIIGVSLMAFQQLSGVNAVMFYAETIFEEA 286
              :|   |   : :|:    | | |     :| || : |   |||||:||| :|: :|||:|
Db        243 AEMKDESRXMXSEKXVSVLELFRSRXYRQPVIIAIVLQLSQQLSGINAVFYYSTSIFEKA 302

Qy        287 KFKDSSLASVVVGVIQVLFTAVAALIMDRAGRRLLLVLSGVVMVFSTSAFGAYFKLTQGG 346
                     |::  ||:  :|| |:  :::||||| | :|                     |
Db        303 GVGQPVYATIGAGVVNTVFTVVSVFVVERAGRRTLHLL---------------------G 341

Qy        347 PGNSSHVAISAPVSAQPVDASVGLAWLAVGSMCLFIAGFAVGWGPIPWLLMSEIFPLHVK 406
               |  :  |:   ::   :|    :::::: ::  |:| | || ||||| :::|:|    :
Db        342 LGGMAGCAVLMTIALALLDQVPWMSYVSIVAIFGFVAFFEVGPGPIPWFIVAELFSQGPR 401

Qy        407 GVATGICVLTNWLMAFLVTKEFSSLMEVLRPYGAFWLASAFCIFSVL------FTLFCVP 460
                |  :   :||   |:|   |  : |:| ||        | :|:||      ||   ||
Db        402 PAAIAVAGFSNWTSNFIVGLLFQYIAELLGPY-------VFIVFAVLLLLFFIFTFLKVP 454

Qy        461 ETKGKTLEQITAHF 474
              ||||:| ::| | |
Db        455 ETKGRTFDEIAAAF 468

RESULT 10
US-09-679-686B-23
; Sequence 23, Application US/09679686B
; Patent No. 6624343
; GENERAL INFORMATION:
;  APPLICANT: Allen, Stephen M.
;  APPLICANT:  Lightner, Jonathan E.
;  APPLICANT:  Rafalski, J. Antoni
;  APPLICANT:  Thorpe, Catherine J.
;  TITLE OF INVENTION: HEXOSE CARRIER PROTEINS
;  FILE REFERENCE: BB1160 US NA
;  CURRENT APPLICATION NUMBER: US/09/679,686B
;  CURRENT FILING DATE:  2003-01-16
;  PRIOR APPLICATION NUMBER: 60/081,131
;  PRIOR FILING DATE: 1998-04-09
;  PRIOR APPLICATION NUMBER: PCT/US99/07561
;  PRIOR FILING DATE: 1999-04-07
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 23
;   LENGTH: 518
;   TYPE: PRT
;   ORGANISM: Medicago truncatula
US-09-679-686B-23

  Query Match             22.3%;  Score 548.5;  DB 4;  Length 518;
  Best Local Similarity   30.4%;  Pred. No. 2e-43;
  Matches  151;  Conservative   88;  Mismatches  190;  Indels   67;  Gaps   16;

Qy         16 PGGSAPRGRRVFLAAFAAALGPLSFGFALGYSSPAI---PSLQRAAPPAPR--------- 63
              ||   |    | :    ||:| | ||: :| |       | |::  |   |
Db         16 PGNLTP---FVTITCIVAAMGGLIFGYDIGISGGVTSMDPFLKKFFPAVYRKKNKDKSTN 72

Qy         64 ----LDDAAASWFGAVVTLGAAAGGVLGGWLVDRAGRKLSLLLCSVPFVAGFAVITAAQD 119
                   |    : | : : | |    ::   :  | |||||:|   : |: |  :   |
Db         73 QYCQYDSQTLTMFTSSLYLAALLSSLVASTITRRFGRKLSMLFGGLLFLVGALINGFANH 132

Qy        120 VWMLLGGRLLTGLACGVASLVAPVYISEIAYPAVRGLLGSCVQLMVVVGILLAYLAGWVL 179
              ||||: ||:| |   | |:   |:|:||:|    || |    || : :|||:| :  :
Db        133 VWMLIVGRILLGFGIGFANQPVPLYLSEMAPYKYRGALNIGFQLSITIGILVANVLNYFF 192

Qy        180 E-----WRW-LAVLGCVPPSLMLLL-MCFMPETPRFLLTQHRRQEAMAALRFLWGSE--- 229
                    | | |:: | : |:|:: :    :|:||  :: :  |  | | |: : | |
Db        193 AKIKGGWGWRLSLGGAMVPALIITIGSLVLPDTPNSMIERGDRDGAKAQLKRIRGIEDVD 252

Qy        230 QGWEDPPIGAEQSFHLA-----LLRQPGIYKP-FIIGVSLMAFQQLSGVNAVMFYAETIF 283
              : : |    :| |  :      ||::   |:|   : | :  ||| :|:| :||||  :|
Db        253 EEFNDLVAASEASMQVENPWRNLLQRK--YRPQLTMAVLIPFFQQFTGINVIMFYAPVLF 310

Qy        284 EEAKFKD--SSLASVVVGVIQVLFTAVAALIMDRAGRRLLLVLSGVVM----VFSTSAFG 337
                  |||  | :::|: ||: |: | |:   :|: ||| | :  |  |    |   :| |
Db        311 NSIGFKDDASLMSAVITGVVNVVATCVSIYGVDKWGRRALFLEGGAQMLICQVAVAAAIG 370

Qy        338 AYFKLTQGGPGNSSHVAISAPVSAQPVDASVGLAWLA---VGSMCLFIAGFAVGWGPIPW 394
              | |  | | |||                      | |   |  :|:::||||  |||: |
Db        371 AKFG-TSGNPGNLPE-------------------WYAIVVVLFICIYVAGFAWSWGPLGW 410
```

```
Qy         395 LLMSEIFPLHVKGVATGICVLTNWLMAFLVTKEFSSLMEVLRPYGAFWLASAFCIFSVLF 454
               |: |||||| ::  |  : |  | |  ||| : |  ::  :: :| |   : | :   ::
Db         411 LVPSEIFPLEIRSAAQSVNVSVNMLFTFLVAQVFLIMLCHMK-FGLFLFFAFFVLVMSIY 469

Qy         455 TLFCVPETKGKTLEQI 470
                 | :|||||  :|::
Db         470 VFFLLPETKGIPIEEM 485

RESULT 11
US-09-679-686B-18
; Sequence 18, Application US/09679686B
; Patent No. 6624343
; GENERAL INFORMATION:
;  APPLICANT: Allen, Stephen M.
;  APPLICANT:  Lightner, Jonathan E.
;  APPLICANT:  Rafalski, J. Antoni
;  APPLICANT:  Thorpe, Catherine J.
;  TITLE OF INVENTION: HEXOSE CARRIER PROTEINS
;  FILE REFERENCE: BB1160 US NA
;  CURRENT APPLICATION NUMBER: US/09/679,686B
;  CURRENT FILING DATE:  2003-01-16
;  PRIOR APPLICATION NUMBER: 60/081,131
;  PRIOR FILING DATE: 1998-04-09
;  PRIOR APPLICATION NUMBER: PCT/US99/07561
;  PRIOR FILING DATE: 1999-04-07
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 18
;   LENGTH: 517
;   TYPE: PRT
;   ORGANISM: Triticum aestivum
US-09-679-686B-18

  Query Match             22.1%;  Score 542.5;  DB 4;  Length 517;
  Best Local Similarity   29.9%;  Pred. No. 7.4e-43;
  Matches  150;  Conservative   88;  Mismatches  202;  Indels   61;  Gaps   14;

Qy          17 GGSAPRGRR---VFLAAFAAALGPLSFGFALGYSSPAI---PSLQRAAPPA--------- 61
               ||    ||    ||     || | | ||: :| |       | |::  |
Db          11 GGKDYPGRLTLFVFFTCVVAATGGLIFGYDIGISGGVTSMNPFLKKFFPEVYDKKQMKGS 70

Qy          62 ----PRLDDAAASWFGAVVTLGAAAGGVLGGWLVDRAGRKLSLLLCSVPFVAGFAVITAA 117
                    : |:     | : : | |         :    ||| |:    : |: | |:  ||
Db          71 ASQYCKYDNQLLQTFTSSLYLAALVSSFFAATVTRVVGRKWSMFTGGLTFLIGAALNGAA 130

Qy         118 QDVWMLLGGRLLTGLACGVASLVAPVYISEIAYPAVRGLLGSCVQLMVVVGILLAYL--- 174
               ::: ||: ||:| |:  | |:   |||:||:|   :||:|    |||: :||| | |
Db         131 ENIAMLIVGRILLGVGVGFANQSVPVYLSEMAPARLRGMLNIGFQLMITIGILAAALINY 190

Qy         175 ------AGWVLEWRWLAVLGCVPPSLMLLLMCFMPETPRFLLTQHRRQEAMAALRFLWGS 228
                     ||:   ||    :  ||  :: |   |:|:||  |: :   : |   |  : ||
Db         191 DTNKIKAGY--GWRISLAIAAVPAGIITLGSFFLPDTPNSLIERGHPEAARRMLNRIRGS 248

Qy         229 E----QGWEDPPIGAEQSFHLALLRQP------GIYKP-FIIGVSLMAFQQLSGVNAVMF 277
               :    : : |  : :|:|    |:: |        |:|   : : :  ||||:|:| :||
Db         249 DVDISEEYADLVVASEES---KLVQHPWRNILQRKYRPQLTMAIMIPFFQQLTGINVIMF 305

Qy         278 YAETIFEEAKFK-DSSLAS-VVVGVIQVLFTAVAALIMDRAGRRLLLVLSGVVMVFSTSA 335
               ||  :||   || |:|| | |: |:: |  | |:   :|| ||| | :  |  |: |
Db         306 YAPVLFETLGFKGDASLMSAVITGLVNVFATLVSVFTVDRLGRRKLFLQGGTQMLLSQLV 365

Qy         336 FGAYFKLTQGGPGNSSHVAISAPVSAQPVDASVGLAWLAVGSMCLFIAGFAVGWGPIPWL 395
                |    :  |  |          |   |     | |   |  :||::||||  |||: ||
Db         366 VGTLIAVKFGTSG----------VGEMP----KGYAAAVVLFICLYVAGFAWSWGPLGWL 411

Qy         396 LMSEIFPLHVKGVATGICVLTNWLMAFLVTKEFSSLMEVLRPYGAFWLASAFCIFSVLFT 455
               : |||||| ::     | |  | |  |:: : | :::  :: :| |:  : : :   :|
Db         412 VPSEIFPLEIRPAGQSINVSVNMLFTFVIAQAFLTMLCHMK-FGLFYFFAGWVVIMTVFI 470

Qy         456 LFCVPETKGKTLEQITAHFEG 476
                  :||||   :|::   ::|
Db         471 ALFLPETKNVPIEEMVLVWKG 491
```

```
RESULT 12
US-09-031-392-6
; Sequence 6, Application US/09031392
; Patent No. 5942398
;  GENERAL INFORMATION:
;    APPLICANT:  Tartaglia, Louis A.
;    APPLICANT:  Weng, Xun
;    TITLE OF INVENTION:  NUCLEIC ACID MOLECULES
;    TITLE OF INVENTION:  ENCODING GLUTEX AND USES THEREOF
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Windows95
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/031,392
;      FILING DATE:  26-FEB-1998
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meiklejohn, Ph.D., Anita L.
;      REGISTRATION NUMBER:  35,283
;      REFERENCE/DOCKET NUMBER:  07334/072001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/542-5070
;      TELEFAX:  617/542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  509 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-031-392-6

  Query Match             21.8%;  Score 536;  DB 2;  Length 509;
  Best Local Similarity   30.5%;  Pred. No. 3e-42;
  Matches  159;  Conservative   86;  Mismatches  172;  Indels  104;  Gaps   16;

Qy          13 LGPPGGSAPRGR---RVFLAAFAAALGPLSFGFA-------------------LGYSSPA 50
               :|   |  |: |    : || |:| || | ||:                    ||   |
Db           8 IGSEDGEPPQQRVTGTLVLAVFSAVLGSLQFGYNIGVINAPQKVIEQSYNETWLGRQGPE 67

Qy          51 IPSLQRAAPPAPRLDDAAASWFGAVVTLGA------AAGGVLGGWLV----DRAGRKLSL 100
                ||   : ||            | : || |      : ||::  :|:       ||| ::
Db          68 GPS---SIPP------------GTLTTLWALSVAIFSVGGMISSFLIGIISQWLGRKRAM 112

Qy         101 LLCSVPFVAG---FAVITAAQDVWMLLGGRLLTGLACGVASLVAPVYISEIAYPAVRGLL 157
               |: :|  | |     :  ||    ||: || | |   |: | : |:|: |||   :|| |
Db         113 LVNNVLAVLGGSLMGLANAAASYEMLILGRFLIGAYSGLTSGLVPMYVGEIAPTHLRGAL 172

Qy         158 GSCVQLMVVVGILLAYLAGWVLE--------WRWLAVLGCVPPSLMLLLMCFMPETPRFL 209
               |:  || :|:|||:| : |  ||        |  |  |  :|  | |:|: | ||:||:|
Db         173 GTLNQLAIVIGILIAQVLG--LESLLGTASLWPLLLGLTVLPALLQLVLLPFCPESPRYL 230

Qy         210 -LTQHRRQEAMAALRFL--WGSEQG----WEDPPIGAEQSFHLALLRQPGI---YKPFII 259
                : |:    |  :|: |  |    |     :|     |:   |:||:  |     :| ||
Db         231 YIIQNLEGPARKSLKRLTGWADVSGVLAELKDEKRKLERERPLSLLQLLGSRTHRQPLII 290

Qy         260 GVSLMAFQQLSGVNAVMFYAETIFEEAKFKDSSLASVVVGVIQVLFTAVAALIMDRAGRR 319
                | |   |||||:||| :|: :||| |     : |::  ||:  :|| |: |:::|||||
Db         291 AVVLQLSQQLSGINAVFYYSTSIFETAGVGQPAYATIGAGVVNTVFTLVSVLLVERAGRR 350

Qy         320 LLLVLSGVVMVFSTSAFGAYFKLTQGGPGNSSHVAISAPVSAQPVDASVGLAWLAVGSMC 379
                | :|                     |       ||   |:   ::    :::::: ::
```

```
Db         351 TLHLL---------------------GLAGMCGCAILMTVALLLLERVPAMSYVSIVAIF 389

Qy         380 LFIAGFAVGWGPIPWLLMSEIFPLHVKGVATGICVLTNWLMAFLVTKEFSSLMEVLRPYG 439
                |:| | :| ||||| :::|:|    :  |  :   :||   |::   |  : | : ||
Db         390 GFVAFFEIGPGPIPWFIVAELFSQGPRPAAMAVAGFSNWTSNFIIGMGFQYVAEAMGPY- 448

Qy         440 AFWLASAFCIFSVL------FTLFCVPETKGKTLEQITAHF 474
                      | :|:||      ||   ||||:|:| :||:| |
Db         449 ------VFLLFAVLLLGFFIFTFLRVPETRGRTFDQISAAF 483


RESULT 13
US-09-299-549-6
; Sequence 6, Application US/09299549
; Patent No. 6136547
;  GENERAL INFORMATION:
;    APPLICANT:  Tartaglia, Louis A.
;    APPLICANT:  Weng, Xun
;    TITLE OF INVENTION:  NUCLEIC ACID MOLECULES ENCODING
;    TITLE OF INVENTION:  GLUTEX AND USES THEREOF
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Windows95
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/299,549
;      FILING DATE:  26-APR-1999
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  09/031,392
;      FILING DATE:  26-FEB-1998
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meiklejohn, Ph.D., Anita L.
;      REGISTRATION NUMBER:  35,283
;      REFERENCE/DOCKET NUMBER:  07334/072002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/542-5070
;      TELEFAX:  617/542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  509 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;   MOLECULE TYPE:  protein
US-09-299-549-6

  Query Match             21.8%;  Score 536;  DB 3;  Length 509;
  Best Local Similarity   30.5%;  Pred. No. 3e-42;
  Matches  159;  Conservative   86;  Mismatches  172;  Indels  104;  Gaps   16;

Qy          13 LGPPGGSAPRGR---RVFLAAFAAALGPLSFGFA-------------------LGYSSPA 50
               :|   |  |: |    : || |:| || | ||:                    ||   |
Db           8 IGSEDGEPPQQRVTGTLVLAVFSAVLGSLQFGYNIGVINAPQKVIEQSYNETWLGRQGPE 67

Qy          51 IPSLQRAAPPAPRLDDAAASWFGAVVTLGA------AAGGVLGGWLV----DRAGRKLSL 100
                ||   : ||            | : || |      : ||::  :|:       ||| ::
Db          68 GPS---SIPP------------GTLTTLWALSVAIFSVGGMISSFLIGIISQWLGRKRAM 112

Qy         101 LLCSVPFVAG---FAVITAAQDVWMLLGGRLLTGLACGVASLVAPVYISEIAYPAVRGLL 157
               |: :|  | |     :  ||    ||: || | |   |: | : |:|: |||   :|| |
Db         113 LVNNVLAVLGGSLMGLANAAASYEMLILGRFLIGAYSGLTSGLVPMYVGEIAPTHLRGAL 172

Qy         158 GSCVQLMVVVGILLAYLAGWVLE--------WRWLAVLGCVPPSLMLLLMCFMPETPRFL 209
               |:  || :|:|||:| : |  ||        |  |  |  :|  | |:|: | ||:||:|
Db         173 GTLNQLAIVIGILIAQVLG--LESLLGTASLWPLLLGLTVLPALLQLVLLPFCPESPRYL 230

Qy         210 -LTQHRRQEAMAALRFL--WGSEQG----WEDPPIGAEQSFHLALLRQPGI---YKPFII 259
                : |:    |  :|: |  |    |     :|     |:   |:||:  |     :| ||
Db         231 YIIQNLEGPARKSLKRLTGWADVSGVLAELKDEKRKLERERPLSLLQLLGSRTHRQPLII 290

Qy         260 GVSLMAFQQLSGVNAVMFYAETIFEEAKFKDSSLASVVVGVIQVLFTAVAALIMDRAGRR 319
                | |   |||||:||| :|: :||| |     : |::  ||:  :|| |: |:::|||||
Db         291 AVVLQLSQQLSGINAVFYYSTSIFETAGVGQPAYATIGAGVVNTVFTLVSVLLVERAGRR 350

Qy         320 LLLVLSGVVMVFSTSAFGAYFKLTQGGPGNSSHVAISAPVSAQPVDASVGLAWLAVGSMC 379
                | :|                     |       ||   |:   ::    :::::: ::
Db         351 TLHLL---------------------GLAGMCGCAILMTVALLLLERVPAMSYVSIVAIF 389

Qy         380 LFIAGFAVGWGPIPWLLMSEIFPLHVKGVATGICVLTNWLMAFLVTKEFSSLMEVLRPYG 439
                |:| | :| ||||| :::|:|    :  |  :   :||   |::   |  : | : ||
Db         390 GFVAFFEIGPGPIPWFIVAELFSQGPRPAAMAVAGFSNWTSNFIIGMGFQYVAEAMGPY- 448

Qy         440 AFWLASAFCIFSVL------FTLFCVPETKGKTLEQITAHF 474
                      | :|:||      ||   ||||:|:| :||:| |
Db         449 ------VFLLFAVLLLGFFIFTFLRVPETRGRTFDQISAAF 483


RESULT 14
US-09-610-417-6
; Sequence 6, Application US/09610417
; Patent No. 6346374
;   GENERAL INFORMATION:
;        APPLICANT: Tartaglia, Louis A.
;                   Weng, Xun
;        TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
;                            GLUTEX AND USES THEREOF
;        NUMBER OF SEQUENCES: 10
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Fish & Richardson P.C.
;             STREET: 225 Franklin Street
;             CITY: Boston
;             STATE: MA
;             COUNTRY: USA
;             ZIP: 02110-2804
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Diskette
;             COMPUTER: IBM Compatible
;             OPERATING SYSTEM: Windows95
;             SOFTWARE: FastSEQ for Windows Version 2.0
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/610,417
;             FILING DATE: 05-Jul-2000
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 09/299,549
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Meiklejohn, Ph.D., Anita L.
;             REGISTRATION NUMBER: 35,283
;             REFERENCE/DOCKET NUMBER: 07334/072002
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 617/542-5070
;             TELEFAX: 617/542-8906
;             TELEX: 200154
;   INFORMATION FOR SEQ ID NO: 6:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 509 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-610-417-6

  Query Match             21.8%;  Score 536;  DB 4;  Length 509;
  Best Local Similarity   30.5%;  Pred. No. 3e-42;
  Matches  159;  Conservative   86;  Mismatches  172;  Indels  104;  Gaps   16;

Qy          13 LGPPGGSAPRGR---RVFLAAFAAALGPLSFGFA-------------------LGYSSPA 50
```

```
              :|   |  |: |    : || |:| || | ||:                    ||   |
Db          8 IGSEDGEPPQQRVTGTLVLAVFSAVLGSLQFGYNIGVINAPQKVIEQSYNETWLGRQGPE 67

Qy         51 IPSLQRAAPPAPRLDDAAASWFGAVVTLGA------AAGGVLGGWLV----DRAGRKLSL 100
               ||   : ||            | : || |      : ||::  :|:       ||| ::
Db         68 GPS---SIPP------------GTLTTLWALSVAIFSVGGMISSFLIGIISQWLGRKRAM 112

Qy        101 LLCSVPFVAG---FAVITAAQDVWMLLGGRLLTGLACGVASLVAPVYISEIAYPAVRGLL 157
              |: :|  | |     :  ||    ||: || | |   |: | : |:|: |||   :|| |
Db        113 LVNNVLAVLGGSLMGLANAAASYEMLILGRFLIGAYSGLTSGLVPMYVGEIAPTHLRGAL 172

Qy        158 GSCVQLMVVVGILLAYLAGWVLE--------WRWLAVLGCVPPSLMLLLMCFMPETPRFL 209
              |:  || :|:|||:| : |  ||        |  |  |  :|  | |:|: | ||:||:|
Db        173 GTLNQLAIVIGILIAQVLG--LESLLGTASLWPLLLGLTVLPALLQLVLLPFCPESPRYL 230

Qy        210 -LTQHRRQEAMAALRFL--WGSEQG----WEDPPIGAEQSFHLALLRQPGI---YKPFII 259
               : |:    |  :|: |  |    |     :|     |:   |:||:  |     :| ||
Db        231 YIIQNLEGPARKSLKRLTGWADVSGVLAELKDEKRKLERERPLSLLQLLGSRTHRQPLII 290

Qy        260 GVSLMAFQQLSGVNAVMFYAETIFEEAKFKDSSLASVVVGVIQVLFTAVAALIMDRAGRR 319
               | |   |||||:||| :|: :||| |     : |::  ||:  :|| |: |:::|||||
Db        291 AVVLQLSQQLSGINAVFYYSTSIFETAGVGQPAYATIGAGVVNTVFTLVSVLLVERAGRR 350

Qy        320 LLLVLSGVVMVFSTSAFGAYFKLTQGGPGNSSHVAISAPVSAQPVDASVGLAWLAVGSMC 379
               | :|                     |       ||   |:   ::    :::::: ::
Db        351 TLHLL---------------------GLAGMCGCAILMTVALLLLERVPAMSYVSIVAIF 389

Qy        380 LFIAGFAVGWGPIPWLLMSEIFPLHVKGVATGICVLTNWLMAFLVTKEFSSLMEVLRPYG 439
               |:| | :| ||||| :::|:|    :  |  :   :||   |::   |  : | : ||
Db        390 GFVAFFEIGPGPIPWFIVAELFSQGPRPAAMAVAGFSNWTSNFIIGMGFQYVAEAMGPY- 448

Qy        440 AFWLASAFCIFSVL------FTLFCVPETKGKTLEQITAHF 474
                     | :|:||      ||   ||||:|:| :||:| |
Db        449 ------VFLLFAVLLLGFFIFTFLRVPETRGRTFDQISAAF 483

RESULT 15
US-09-679-686B-22
; Sequence 22, Application US/09679686B
; Patent No. 6624343
; GENERAL INFORMATION:
;  APPLICANT: Allen, Stephen M.
;  APPLICANT:  Lightner, Jonathan E.
;  APPLICANT:  Rafalski, J. Antoni
;  APPLICANT:  Thorpe, Catherine J.
;  TITLE OF INVENTION: HEXOSE CARRIER PROTEINS
;  FILE REFERENCE: BB1160 US NA
;  CURRENT APPLICATION NUMBER: US/09/679,686B
;  CURRENT FILING DATE:  2003-01-16
;  PRIOR APPLICATION NUMBER: 60/081,131
;  PRIOR FILING DATE: 1998-04-09
;  PRIOR APPLICATION NUMBER: PCT/US99/07561
;  PRIOR FILING DATE: 1999-04-07
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 22
;   LENGTH: 514
;   TYPE: PRT
;   ORGANISM: Arabidopsis thaliana
US-09-679-686B-22

  Query Match             21.8%;  Score 535;  DB 4;  Length 514;
  Best Local Similarity   29.7%;  Pred. No. 3.8e-42;
  Matches  146;  Conservative   92;  Mismatches  191;  Indels   62;  Gaps   13;

Qy         25 RVFLAAFAAALGPLSFGFALGYSSPAI---PSLQRAAP------------PAPRLDDAAA 69
              :||:  |  | | | ||: || |       | |:   |               | |
Db         22 KVFVTCFIGAFGGLIFGYDLGISGGVTSMEPFLEEFFPYVYKKMKSAHENEYCRFDSQLL 81

Qy         70 SWFGAVVTLGAAAGGVLGGWLVDRAGRKLSLLLCSVPFVAGFAVITAAQDVWMLLGGRLL 129
              : | : : : |    :    :    ||| |: |    |  | |    ||:: ||| ||:|
Db         82 TLFTSSLYVAALVSSLFASTITRVFGRKWSMFLGGFTFFIGSAFNGFAQNIAMLLIGRIL 141
```

```
Qy        130 TGLACGVASLVAPVYISEIAYPAVRGLLGSCVQLMVVVGILLAYLAGWV-------LEWR 182
               |   | |:   |||:||:| | :||   :  |: :: ||::| :  :        : ||
Db        142 LGFGVGFANQSVPVYLSEMAPPNLRGAFNNGFQVAIIFGIVVATIINYFTAQMKGNIGWR 201

Qy        183 WLAVLGCVPPSLMLLLMCFMPETPRFLLTQHRRQEAMAALRFLWGS---EQGWEDPPIGA 239
                  | |||  ::::    :|:||  |: :   :||   |: : |:   :: ::|    :
Db        202 ISLGLACVPAVMIMIGALILPDTPNSLIERGYTEEAKEMLQSIRGTNEVDEEFQDLIDAS 261

Qy        240 EQS-------FHLALLRQPGIYKPFIIGVSLMA-FQQLSGVNAVMFYAETIFEEAKF--K 289
              |:|        :: | |    |:| :|    :  ||||:|:| : |||  :|:   |  |
Db        262 EESKQVKHPWKNIMLPR----YRPQLIMTCFIPFFQQLTGINVITFYAPVLFQTLGFGSK 317

Qy        290 DSSLASVVVGVIQVLFTAVAALIMDRAGRRLLLVLSGVVMVFSTSAFGAYF--KLTQGGP 347
               | |:::| |:|::| | |:   :|| |||:| :  |: |: |  | ||    |    |
Db        318 ASLLSAMVTGIIELLCTFVSVFTVDRFGRRILFLQGGIQMLVSQIAIGAMIGVKFGVAGT 377

Qy        348 GNSSHVAISAPVSAQPVDASVGLAWLAVGSMCLFIAGFAVGWGPIPWLLMSEIFPLHVKG 407
              ||               ||:     | |  :|:::||||  |||: ||: ||| || ::
Db        378 GNIG-----------KSDAN-----LIVALICIYVAGFAWSWGPLGWLVPSEISPLEIRS 421

Qy        408 VATGICVLTNWLMAFLVTKEFSSLMEVLRPYGAFWLASAFCIFSVLFTLFCVPETKGKTL 467
               |  | |  |    ||| : | :::  :: :| |:  : | :   :|    :||||   :
Db        422 AAQAINVSVNMFFTFLVAQLFLTMLCHMK-FGLFFFFAFFVVIMTIFIYLMLPETKNVPI 480

Qy        468 EQIT----AHF 474
              |::     ||:
Db        481 EEMNRVWKAHW 491

Search completed: September 27, 2004, 19:30:47
Job time : 36 secs
```

```
                         GenCore version 5.1.6
                Copyright (c) 1993 - 2004  Compugen Ltd.

OM protein - protein search, using sw model

Run on:        September 27, 2004, 19:08:41 ; Search time 126 Seconds
                                          (without alignments)
                                          1069.644 Million cell updates/sec

Title:         US-09-886-954A-1
Perfect score: 2457
Sequence:      1 MTPEDPEETQPLLGPPGGSA..........CVPETKGKTLEQITAHFEGR 477

Scoring table: BLOSUM62
               Gapop 10.0 , Gapext 0.5

Searched:      1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters:      1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       A_Geneseq_29Jan04:*
                1:  geneseqp1980s:*
                2:  geneseqp1990s:*
                3:  geneseqp2000s:*
                4:  geneseqp2001s:*
                5:  geneseqp2002s:*
                6:  geneseqp2003as:*
                7:  geneseqp2003bs:*
                8:  geneseqp2004s:*

     Pred. No. is the number of results predicted by chance to have a
     score greater than or equal to the score of the result being printed,
     and is derived by analysis of the total score distribution.

                              SUMMARIES
                     %
Result            Query
  No.    Score    Match Length DB  ID                       Description
-----------------------------------------------------------------------------
   1      2457    100.0    477  4  AAE04888                 Aae04888 Human tra
   2      2457    100.0    477  6  ABP58364                 Abp58364 Human sol
   3      2448     99.6    477  4  AAB66932                 Aab66932 Human GLU
   4    2224.5     90.5    478  4  AAB66939                 Aab66939 GLUTX1 co
   5    2162.5     88.0    478  4  AAB66933                 Aab66933 Rat GLUTX
   6      2150     87.5    477  4  AAB66934                 Aab66934 Murine GL
   7      1491     60.7    326  5  ABB89717                 Abb89717 Human pol
   8      1135     46.2    262  4  AAE06579                 Aae06579 Human pro
   9      1131     46.0    248  5  ABU11283                 Abu11283 cDNA enco
  10     969.5     39.5    503  4  AAB66938                 Aab66938 Rat GLUTX
  11       953     38.8    507  4  AAB66941                 Aab66941 GLUTX3 co
  12       948     38.6    507  4  AAB66937                 Aab66937 Human GLU
  13       948     38.6    507  4  AAM93417                 Aam93417 Human pol
  14       948     38.6    507  6  ADA84077                 Ada84077 Human SLC
  15       948     38.6    507  6  ABP58365                 Abp58365 Human sol
  16       948     38.6    507  6  ABP58363                 Abp58363 Human sol
  17       894     36.4    445  4  AAM93364                 Aam93364 Human pol
  18     788.5     32.1    489  4  ABB63511                 Abb63511 Drosophil
  19       788     32.1    539  4  ABB65350                 Abb65350 Drosophil
  20     764.5     31.1    433  4  ABB63758                 Abb63758 Drosophil
  21       727     29.6    465  4  ABB62408                 Abb62408 Drosophil
  22       705     28.7    497  4  ABB57922                 Abb57922 Drosophil
  23       670     27.3    447  3  AAG13749                 Aag13749 Arabidops
  24       670     27.3    463  3  AAG13748                 Aag13748 Arabidops
  25     669.5     27.2    437  3  AAG13750                 Aag13750 Arabidops
  26     666.5     27.1    444  4  ABB58057                 Abb58057 Drosophil
  27       662     26.9    447  3  AAG20977                 Aag20977 Arabidops
  28       662     26.9    463  3  AAG20976                 Aag20976 Arabidops
  29     661.5     26.9    437  3  AAG20978                 Aag20978 Arabidops
  30     652.5     26.6    448  3  AAG38870                 Aag38870 Arabidops
  31     652.5     26.6    464  3  AAG38869                 Aag38869 Arabidops
  32       652     26.5    438  3  AAG38871                 Aag38871 Arabidops
  33       647     26.3    491  4  ABB65373                 Abb65373 Drosophil
  34       647     26.3    491  4  ABB57923                 Abb57923 Drosophil
  35       615     25.0    488  6  ABP98504                 Abp98504 PFAM cons
  36       615     25.0    488  6  AAE35306                 Aae35306 Human sug
  37       615     25.0    488  7  ADD22918                 Add22918 Human sug
  38     613.5     25.0    487  6  ABG73334                 Abg73334 Consensus
  39       603     24.5    472  5  ABP52164                 Abp52164 E. coli a
  40       603     24.5    472  7  ADD37502                 Add37502 E. coli a
  41     598.5     24.4    465  6  ABU28442                 Abu28442 Protein e
  42     591.5     24.1    464  5  ABP52163                 Abp52163 E. coli g
  43     591.5     24.1    464  6  ABU14994                 Abu14994 Protein e
  44     591.5     24.1    464  7  ADD37501                 Add37501 E. coli g
  45     585.5     23.8    453  6  ABU31194                 Abu31194 Protein e

                              ALIGNMENTS

RESULT 1
AAE04888
ID   AAE04888 standard; protein; 477 AA.
XX
AC   AAE04888;
XX
DT   10-SEP-2001  (first entry)
XX
DE   Human transporter and ion channel-1 (TRICH-1) protein.
XX
KW   Human; transporter and ion channel-1; TRICH-1; vaccine; cystic fibrosis;
KW   gene therapy; amyotrophic lateral sclerosis; amnesia; muscular dystrophy;
KW   hypertension; angina; neurological disorder; asthma; bipolar disorder;
KW   dementia; depression; Alzheimer's disease; epilepsy; mood; arrhythmia;
KW   Pick's disease; ischaemic cerebrovascular disease; AIDS; anxiety; stroke;
KW   Huntington's disease; Parkinson's disease; cerebral neoplasm; allergy;
KW   demyelinating disease; mental disorder; Schizophrenia; polymyositis;
KW   muscle disorder; cardiomyopathy; cataract; myocarditis; Grave's disease;
KW   dermatomyositis; diabetes mellitus; immunological disorder; psoriasis;
KW   rheumatoid arthritis; Sjogren's syndrome; systemic lupus erythematosus;
KW   sickle cell anaemia; Wilson's disease; infertility; Cushing's disease;
KW   scleroderma; pulmonary artery stenosis; nootropic; Addison's disease;
KW   malabsorption syndrome; hypercholesterolaemia; cancer.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Domain          29..474
FT                   /note= "Sugar transporter domain"
FT   Domain          259..279
FT                   /label= Transmembrane_domain
FT   Domain          293..313
FT                   /label= Transmembrane_domain
FT   Domain          320..339
FT                   /label= Transmembrane_domain
FT   Domain          438..457
FT                   /label= Transmembrane_domain
XX
PN   WO200146258-A2.
XX
PD   28-JUN-2001.
XX
PF   22-DEC-2000; 2000WO-US035095.
XX
PR   23-DEC-1999;   99US-0172000P.
PR   14-JAN-2000; 2000US-0176083P.
PR   21-JAN-2000; 2000US-0177332P.
PR   28-JAN-2000; 2000US-0178572P.
```

```
PR   02-FEB-2000; 2000US-0179758P.
PR   10-FEB-2000; 2000US-0181625P.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Baughn MR,  Burford N,  Au-Young J,  Lu DAM,  Yang J,  Reddy R;
PI   Lal P,  Hillman JL,  Azimzai Y,  Yue H,  Nguyen DB,  Yao MG,  Gandhi AR;
PI   Tang YT,  Khan FA;
XX
DR   WPI; 2001-418042/44.
DR   N-PSDB; AAD09552.
XX
PT   Novel human transporter and ion channel proteins useful for treating and
PT   preventing transport, neurological, muscle and immunological disorders.
XX
PS   Claim 1; Page 112-113; 160pp; English.
XX
CC   The present sequence is transporter and ion channel-1 (TRICH-1) protein.
CC   TRICH is used as vaccine. TRICH is useful for treating a disease or
CC   condition associated with decreased expression of functional TRICH, such
CC   as transport disorder including amyotrophic lateral sclerosis, cystic
CC   fibrosis, Becker's muscular dystrophy, Charcot-Marie Tooth disease,
CC   Duchenne muscular dystrophy, angina and hypertension, neurological
CC   disorders including Alzheimer's disease, amnesia, bipolar disorder,
CC   dementia, depression, epilepsy, ischaemic cerebrovascular disease,
CC   stroke, cerebral neoplasms, Pick's disease, Huntington's disease and
CC   Parkinson's disease, demyelinating diseases, mental disorders including
CC   mood, anxiety, Schizophrenia and seasonal affective disorder, muscle
CC   disorder including cardiomyopathy, myocarditis, polymyositis,
CC   dermatomyositis, arrhythmias and asthma and immunological disorders
CC   including AIDS, adult respiratory distress syndrome (ARDS), allergies,
CC   anaemia, diabetes mellitus, rheumatoid arthritis, scleroderma, Sjogren's
CC   syndrome, systemic lupus erythematosus and other diseases including
CC   sickle cell anaemia, Wilson's disease, cataracts, infertility, pulmonary
CC   artery stenosis, Grave's disease, Cushing's disease, Addison's disease,
CC   glucose-galactose malabsorption syndrome, hypercholesterolaemia, cancers
CC   psoriasis and viral, bacterial, fungal, helminthic and protozoal
CC   infections. TRICH DNA is useful in gene therapy and in diagnostic
CC   purposes
XX
SQ   Sequence 477 AA;

  Query Match             100.0%;  Score 2457;  DB 4;  Length 477;
  Best Local Similarity   100.0%;  Pred. No. 5.1e-227;
  Matches  477;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MTPEDPEETQPLLGPPGGSAPRGRRVFLAAFAAALGPLSFGFALGYSSPAIPSLQRAAPP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MTPEDPEETQPLLGPPGGSAPRGRRVFLAAFAAALGPLSFGFALGYSSPAIPSLQRAAPP 60

Qy         61 APRLDDAAASWFGAVVTLGAAAGGVLGGWLVDRAGRKLSLLLCSVPFVAGFAVITAAQDV 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 APRLDDAAASWFGAVVTLGAAAGGVLGGWLVDRAGRKLSLLLCSVPFVAGFAVITAAQDV 120

Qy        121 WMLLGGRLLTGLACGVASLVAPVYISEIAYPAVRGLLGSCVQLMVVVGILLAYLAGWVLE 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 WMLLGGRLLTGLACGVASLVAPVYISEIAYPAVRGLLGSCVQLMVVVGILLAYLAGWVLE 180

Qy        181 WRWLAVLGCVPPSLMLLLMCFMPETPRFLLTQHRRQEAMAALRFLWGSEQGWEDPPIGAE 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 WRWLAVLGCVPPSLMLLLMCFMPETPRFLLTQHRRQEAMAALRFLWGSEQGWEDPPIGAE 240

Qy        241 QSFHLALLRQPGIYKPFIIGVSLMAFQQLSGVNAVMFYAETIFEEAKFKDSSLASVVVGV 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QSFHLALLRQPGIYKPFIIGVSLMAFQQLSGVNAVMFYAETIFEEAKFKDSSLASVVVGV 300

Qy        301 IQVLFTAVAALIMDRAGRRLLLVLSGVVMVFSTSAFGAYFKLTQGGPGNSSHVAISAPVS 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 IQVLFTAVAALIMDRAGRRLLLVLSGVVMVFSTSAFGAYFKLTQGGPGNSSHVAISAPVS 360

Qy        361 AQPVDASVGLAWLAVGSMCLFIAGFAVGWGPIPWLLMSEIFPLHVKGVATGICVLTNWLM 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AQPVDASVGLAWLAVGSMCLFIAGFAVGWGPIPWLLMSEIFPLHVKGVATGICVLTNWLM 420
```

```
Qy        421 AFLVTKEFSSLMEVLRPYGAFWLASAFCIFSVLFTLFCVPETKGKTLEQITAHFEGR 477
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AFLVTKEFSSLMEVLRPYGAFWLASAFCIFSVLFTLFCVPETKGKTLEQITAHFEGR 477


RESULT 2
ABP58364
ID   ABP58364 standard; protein; 477 AA.
XX
AC   ABP58364;
XX
DT   07-APR-2003  (first entry)
XX
DE   Human solute carrier type 2A polypeptide 12735153 and 7657681.
XX
KW   Human; solute carrier type 2A; SLC2A; glucose transporter; p53; cancer;
KW   cytostatic; gene therapy.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Domain          27. .49
FT                   /note= "transmembrane domain"
FT   Domain          29. .474
FT                   /note= "sugar transporter domain"
FT   Domain          69. .91
FT                   /note= "transmembrane domain"
FT   Domain          98. .117
FT                   /note= "transmembrane domain"
FT   Domain          127. .149
FT                   /note= "transmembrane domain"
FT   Domain          156. .178
FT                   /note= "transmembrane domain"
FT   Domain          183. .205
FT                   /note= "transmembrane domain"
FT   Domain          257. .279
FT                   /note= "transmembrane domain"
FT   Domain          294. .313
FT                   /note= "transmembrane domain"
FT   Domain          320. .342
FT                   /note= "transmembrane domain"
FT   Domain          368. .390
FT                   /note= "transmembrane domain"
FT   Domain          403. .425
FT                   /note= "transmembrane domain"
FT   Domain          440. .462
FT                   /note= "transmembrane domain"
XX
PN   WO200298467-A1.
XX
PD   12-DEC-2002.
XX
PF   03-JUN-2002; 2002WO-US017419.
XX
PR   05-JUN-2001; 2001US-0296076P.
PR   10-OCT-2001; 2001US-0328605P.
PR   15-FEB-2002; 2002US-0357253P.
XX
PA   (EXEL-) EXELIXIS INC.
XX
PI   Friedman L,  Plowman GD,  Belvin M,  Francis-Lang H,  Li D,  Funke RP;
XX
DR   WPI; 2003-201283/19.
DR   N-PSDB; ABZ24792, ABZ24793, ABZ24794.
XX
PT   Identifying candidate p53 pathway modulating agent as therapeutic target
PT   for disorders of defective p53 function e.g. cancer, by assaying purified
PT   solute carrier family 2 (SLC2A) polypeptide or nucleic acid with a test
PT   agent.
XX
PS   Claim 13; Page 50-52; 58pp; English.
XX
```

```
CC   The present sequence is that of human solute carrier type 2A (SLC2A)
CC   polypeptides 12735153 and 7657681. In the present invention, genetic
CC   screens were designed to identify modifiers of the p53 pathway in
CC   Drosophila in which p53 was overexpressed. Human orthologues
CC   (polynucleotides and polypeptides) of one such modifier were then
CC   identified, including the present polypeptides. SLC2As are glucose
CC   transporter proteins with sugar transporter domains. They are attractive
CC   drug targets for treatment of pathologies associated with a defective p53
CC   signalling pathway, such as cancer. The invention provides in vitro and
CC   in vivo methods of assessing SLC2A function. Modulation of an SLC2A or
CC   its binding partners is useful for understanding the association of the
CC   p53 pathway and its members in normal and disease conditions and for
CC   developing diagnostic and therapeutic modalities of p53-related
CC   pathologies. SLC2A-modulating agents that act by inhibiting or enhancing
CC   SLC2A expression, directly or indirectly, e.g. by affecting an SLC2A
CC   function such as transport or binding activity, can be identified using
CC   methods provided. Modulators include small molecules, nucleic acids,
CC   antibodies, antisense oligonucleotides and phosphothioate morpholino
CC   oligomers (claimed)
XX
SQ   Sequence 477 AA;

  Query Match             100.0%;  Score 2457;  DB 6;  Length 477;
  Best Local Similarity   100.0%;  Pred. No. 5.1e-227;
  Matches  477;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MTPEDPEETQPLLGPPGGSAPRGRRVFLAAFAAALGPLSFGFALGYSSPAIPSLQRAAPP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MTPEDPEETQPLLGPPGGSAPRGRRVFLAAFAAALGPLSFGFALGYSSPAIPSLQRAAPP 60

Qy         61 APRLDDAAASWFGAVVTLGAAAGGVLGGWLVDRAGRKLSLLLCSVPFVAGFAVITAAQDV 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 APRLDDAAASWFGAVVTLGAAAGGVLGGWLVDRAGRKLSLLLCSVPFVAGFAVITAAQDV 120

Qy        121 WMLLGGRLLTGLACGVASLVAPVYISEIAYPAVRGLLGSCVQLMVVVGILLAYLAGWVLE 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 WMLLGGRLLTGLACGVASLVAPVYISEIAYPAVRGLLGSCVQLMVVVGILLAYLAGWVLE 180

Qy        181 WRWLAVLGCVPPSLMLLLMCFMPETPRFLLTQHRRQEAMAALRFLWGSEQGWEDPPIGAE 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 WRWLAVLGCVPPSLMLLLMCFMPETPRFLLTQHRRQEAMAALRFLWGSEQGWEDPPIGAE 240

Qy        241 QSFHLALLRQPGIYKPFIIGVSLMAFQQLSGVNAVMFYAETIFEEAKFKDSSLASVVVGV 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QSFHLALLRQPGIYKPFIIGVSLMAFQQLSGVNAVMFYAETIFEEAKFKDSSLASVVVGV 300

Qy        301 IQVLFTAVAALIMDRAGRRLLLVLSGVVMVFSTSAFGAYFKLTQGGPGNSSHVAISAPVS 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 IQVLFTAVAALIMDRAGRRLLLVLSGVVMVFSTSAFGAYFKLTQGGPGNSSHVAISAPVS 360

Qy        361 AQPVDASVGLAWLAVGSMCLFIAGFAVGWGPIPWLLMSEIFPLHVKGVATGICVLTNWLM 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AQPVDASVGLAWLAVGSMCLFIAGFAVGWGPIPWLLMSEIFPLHVKGVATGICVLTNWLM 420

Qy        421 AFLVTKEFSSLMEVLRPYGAFWLASAFCIFSVLFTLFCVPETKGKTLEQITAHFEGR 477
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AFLVTKEFSSLMEVLRPYGAFWLASAFCIFSVLFTLFCVPETKGKTLEQITAHFEGR 477


RESULT 3
AAB66932
ID   AAB66932 standard; protein; 477 AA.
XX
AC   AAB66932;
XX
DT   17-APR-2001  (first entry)
XX
DE   Human GLUTX1.
XX
KW   Human; GLUTX; gene therapy; vaccine; hexose transport modulator;
KW   hexose transport disorder; ischaemia; diabetes; hyperglycaemia;
KW   hypoglycaemia; glucose metabolism disorder; neurodegenerative disease.
XX
```

```
OS   Homo sapiens.
XX
PN   WO200104145-A2.
XX
PD   18-JAN-2001.
XX
PF   14-JUL-2000; 2000WO-IB001042.
XX
PR   14-JUL-1999;   99US-0143907P.
PR   27-AUG-1999;   99US-0151140P.
PR   23-FEB-2000; 2000US-0184285P.
PR   13-JUL-2000; 2000US-00616132.
XX
PA   (UYLA-) UNIV LAUSANNE.
XX
PI   Thorens B,  Ibberson M,  Uldry M;
XX
DR   WPI; 2001-112615/12.
DR   N-PSDB; AAF55865.
XX
PT   Nucleic acids encoding GLUTX glucose transporter proteins, useful in the
PT   prevention, diagnosis and treatment of hexose transport disorders, e.g.
PT   ischemia and diabetes.
XX
PS   Claim 11; Page 70-71; 124pp; English.
XX
CC   The present invention relates to GLUTX proteins (AAF55865-AAF55871 and
CC   AAB66932-AAB66941). The GLUTX proteins are related to the facultative
CC   glucose carriers GLUT1-GLUT5 and have hexose binding and/or transport
CC   function. The GLUTX proteins may be used in the diagnosis, prevention and
CC   treatment of hexose transport disorders such as ischaemia, diabetes,
CC   hyperglycaemia, hypoglycaemia, a glucose metabolism disorder and/or a
CC   neurodegenerative disease. The present sequence is human GLUTX1
XX
SQ   Sequence 477 AA;

  Query Match             99.6%;  Score 2448;  DB 4;  Length 477;
  Best Local Similarity   99.6%;  Pred. No. 3.7e-226;
  Matches  475;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 MTPEDPEETQPLLGPPGGSAPRGRRVFLAAFAAALGPLSFGFALGYSSPAIPSLQRAAPP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MTPEDPEETQPLLGPPGGSAPRGRRVFLAAFAAALGPLSFGFALGYSSPAIPSLQRAAPP 60

Qy         61 APRLDDAAASWFGAVVTLGAAAGGVLGGWLVDRAGRKLSLLLCSVPFVAGFAVITAAQDV 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 APRLDDAAASWFGAVVTLGAAAGGVLGGWLVDRAGRKLSLLLCSVPFVAGFAVITAAQDV 120

Qy        121 WMLLGGRLLTGLACGVASLVAPVYISEIAYPAVRGLLGSCVQLMVVVGILLAYLAGWVLE 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 WMLLGGRLLTGLACGVASLVAPVYISEIAYPAVRGLLGSCVQLMVVVGILLAYLAGWVLE 180

Qy        181 WRWLAVLGCVPPSLMLLLMCFMPETPRFLLTQHRRQEAMAALRFLWGSEQGWEDPPIGAE 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 WRWLAVLGCVPPSLMLLLMCFMPETPRFLLTQHRRQEAMAALRFLWGSEQGWEDPPIGAE 240

Qy        241 QSFHLALLRQPGIYKPFIIGVSLMAFQQLSGVNAVMFYAETIFEEAKFKDSSLASVVVGV 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QSFHLALLRQPGIYKPFIIGVSLMAFQQLSGVNAVMFYAETIFEEAKFKDSSLASVVVGV 300

Qy        301 IQVLFTAVAALIMDRAGRRLLLVLSGVVMVFSTSAFGAYFKLTQGGPGNSSHVAISAPVS 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 IQVLFTAVAALIMDRAGRRLLLVLSGVVMVFSTSAFGAYFKLTQGGPGNSSHVAISAPVS 360

Qy        361 AQPVDASVGLAWLAVGSMCLFIAGFAVGWGPIPWLLMSEIFPLHVKGVATGICVLTNWLM 420
              ||||||||||||||||:|||||||||||||||||||||||||||||||||||||||||||
Db        361 AQPVDASVGLAWLAVGNMCLFIAGFAVGWGPIPWLLMSEIFPLHVKGVATGICVLTNWLM 420

Qy        421 AFLVTKEFSSLMEVLRPYGAFWLASAFCIFSVLFTLFCVPETKGKTLEQITAHFEGR 477
              ||||||||||||||||||||||||||||||||||||||||| |||||||||||||||
Db        421 AFLVTKEFSSLMEVLRPYGAFWLASAFCIFSVLFTLFCVPEIKGKTLEQITAHFEGR 477
```

```
RESULT 4
AAB66939
ID   AAB66939 standard; protein; 478 AA.
XX
AC   AAB66939;
XX
DT   17-APR-2001  (first entry)
XX
DE   GLUTX1 consensus sequence.
XX
KW   GLUTX; gene therapy; vaccine; hexose transport modulator; human; rat;
KW   hexose transport disorder; ischaemia; diabetes; hyperglycaemia; murine;
KW   hypoglycaemia; glucose metabolism disorder; neurodegenerative disease.
XX
OS   Homo sapiens.
OS   Mus sp.
OS   Rattus sp.
XX
PN   WO200104145-A2.
XX
PD   18-JAN-2001.
XX
PF   14-JUL-2000; 2000WO-IB001042.
XX
PR   14-JUL-1999;   99US-0143907P.
PR   27-AUG-1999;   99US-0151140P.
PR   23-FEB-2000; 2000US-0184285P.
PR   13-JUL-2000; 2000US-00616132.
XX
PA   (UYLA-) UNIV LAUSANNE.
XX
PI   Thorens B,  Ibberson M,  Uldry M;
XX
DR   WPI; 2001-112615/12.
XX
PT   Nucleic acids encoding GLUTX glucose transporter proteins, useful in the
PT   prevention, diagnosis and treatment of hexose transport disorders, e.g.
PT   ischemia and diabetes.
XX
PS   Claim 11; Page 74-75; 124pp; English.
XX
CC   The present invention relates to GLUTX proteins (AAF55865-AAF55871 and
CC   AAB66932-AAB66941). The GLUTX proteins are related to the facultative
CC   glucose carriers GLUT1-GLUT5 and have hexose binding and/or transport
CC   function. The GLUTX proteins may be used in the diagnosis, prevention and
CC   treatment of hexose transport disorders such as ischaemia, diabetes,
CC   hyperglycaemia, hypoglycaemia, a glucose metabolism disorder and/or a
CC   neurodegenerative disease. The present sequence is a consensus sequence
CC   for GLUTX1
XX
SQ   Sequence 478 AA;

  Query Match             90.5%;  Score 2224.5;  DB 4;  Length 478;
  Best Local Similarity   88.9%;  Pred. No. 1.1e-204;
  Matches  425;  Conservative   23;  Mismatches   29;  Indels    1;  Gaps    1;

Qy          1 MTPEDPEETQPLLGPPGGSAPRGRRVFLAAFAAALGPLSFGFALGYSSPAIPSLQRAAPP 60
              |:||||:|||||| |||  ||||||||||||||||||||||||| |||||||||:| |||
Db          1 MSPEDPQETQPLLRPPGARAPRGRRVFLAAFAAALGPLSFGFALBYSSPAIPSLRRTAPP 60

Qy         61 APRLDDAAASWFGAVVTLGAAAGGVLGGWLVDRAGRKLSLLLCSVPFVAGFAVITAAQDV 120
              | || |||||||||||||||||||||||||:||||||||||||:|||| ||||||||:||
Db         61 ALRLGDAAASWFGAVVTLGAAAGGVLGGWLLDRAGRKLSLLLCTVPFVTGFAVITAARDV 120

Qy        121 WMLLGGRLLTGLACGVASLVAPVYISEIAYPAVRGLLGSCVQLMVVVGILLAYLAGWVLE 180
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||:||||||
Db        121 WMLLGGRLLTGLACGVASLVAPVYISEIAYPAVRGLLGSCVQLMVVTGILLAYVAGWVLE 180

Qy        181 WRWLAVLGCVPPSLMLLLMCFMPETPRFLLTQHRRQEAMAALRFLWGSEQGWEDPPIGAE 240
              ||||||||||||:|||||||:||||||||||||: ||||||||||||||:|||:||:|||
Db        181 WRWLAVLGCVPPTLMLLLMCYMPETPRFLLTQHQYQEAMAALRFLWGSEEGWEEPPVGAE 240

Qy        241 -QSFHLALLRQPGIYKPFIIGVSLMAFQQLSGVNAVMFYAETIFEEAKFKDSSLASVVVG 299
               | | |||||:|||||| |||:||| |||||||||:|||| |||||||||||||||| ||
Db        241 HQGFQLALLRRPGIYKPLIIGISLMVFQQLSGVNAIMFYANTIFEEAKFKDSSLASVTVG 300

Qy        300 VIQVLFTAVAALIMDRAGRRLLLVLSGVVMVFSTSAFGAYFKLTQGGPGNSSHVAISAPV 359
              :|||||||||||||||||||||| ||||:|||| |||| |||||| || ||||| |  |:
Db        301 IIQVLFTAVAALIMDRAGRRLLLALSGVIMVFSMSAFGTYFKLTQSGPSNSSHVGILVPI 360

Qy        360 SAQPVDASVGLAWLAVGSMCLFIAGFAVGWGPIPWLLMSEIFPLHVKGVATGICVLTNWL 419
              ||:||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 SAEPVDVSVGLAWLAVGSMCLFIAGFAVGWGPIPWLLMSEIFPLHVKGVATGICVLTNWF 420

Qy        420 MAFLVTKEFSSLMEVLRPYGAFWLASAFCIFSVLFTLFCVPETKGKTLEQITAHFEGR 477
              |||||||||:|||||||||||||| :|||| |||||| |||||||:||||||||||||
Db        421 MAFLVTKEFNSLMEVLRPYGAFWLTAAFCILSVLFTLTCVPETKGRTLEQITAHFEGR 478


RESULT 5
AAB66933
ID   AAB66933 standard; protein; 478 AA.
XX
AC   AAB66933;
XX
DT   17-APR-2001  (first entry)
XX
DE   Rat GLUTX1.
XX
KW   Rat; GLUTX; gene therapy; vaccine; hexose transport modulator;
KW   hexose transport disorder; ischaemia; diabetes; hyperglycaemia;
KW   hypoglycaemia; glucose metabolism disorder; neurodegenerative disease.
XX
OS   Rattus sp.
XX
PN   WO200104145-A2.
XX
PD   18-JAN-2001.
XX
PF   14-JUL-2000; 2000WO-IB001042.
XX
PR   14-JUL-1999;   99US-0143907P.
PR   27-AUG-1999;   99US-0151140P.
PR   23-FEB-2000; 2000US-0184285P.
PR   13-JUL-2000; 2000US-00616132.
XX
PA   (UYLA-) UNIV LAUSANNE.
XX
PI   Thorens B,  Ibberson M,  Uldry M;
XX
DR   WPI; 2001-112615/12.
DR   N-PSDB; AAF55866.
XX
PT   Nucleic acids encoding GLUTX glucose transporter proteins, useful in the
PT   prevention, diagnosis and treatment of hexose transport disorders, e.g.
PT   ischemia and diabetes.
XX
PS   Claim 11; Page 71-72; 124pp; English.
XX
CC   The present invention relates to GLUTX proteins (AAF55865-AAF55871 and
CC   AAB66932-AAB66941). The GLUTX proteins are related to the facultative
CC   glucose carriers GLUT1-GLUT5 and have hexose binding and/or transport
CC   function. The GLUTX proteins may be used in the diagnosis, prevention and
CC   treatment of hexose transport disorders such as ischaemia, diabetes,
CC   hyperglycaemia, hypoglycaemia, a glucose metabolism disorder and/or a
CC   neurodegenerative disease. The present sequence is rat GLUTX1
XX
SQ   Sequence 478 AA;

  Query Match             88.0%;  Score 2162.5;  DB 4;  Length 478;
  Best Local Similarity   85.4%;  Pred. No. 9.9e-199;
  Matches  408;  Conservative   33;  Mismatches   36;  Indels    1;  Gaps    1;

Qy          1 MTPEDPEETQPLLGPPGGSAPRGRRVFLAAFAAALGPLSFGFALGYSSPAIPSLQRAAPP 60
              |:||||:||||||  ||  || ||||||| ||||||||||||||||||||||||:| |||
Db          1 MSPEDPQETQPLLRSPGARAPGGRRVFLATFAAALGPLSFGFALGYSSPAIPSLRRTAPP 60
```

```
Qy          61 APRLDDAAASWFGAVVTLGAAAGGVLGGWLVDRAGRKLSLLLCSVPFVAGFAVITAAQDV 120
               | || | |||||||||||||||||||||||:||||||||||||:|||| ||||||||:||
Db          61 ALRLGDTAASWFGAVVTLGAAAGGVLGGWLLDRAGRKLSLLLCTVPFVTGFAVITAARDV 120

Qy         121 WMLLGGRLLTGLACGVASLVAPVYISEIAYPAVRGLLGSCVQLMVVVGILLAYLAGWVLE 180
               |||||||||||||||||||||||||||||||||||||||||||||| ||||||:||||||
Db         121 WMLLGGRLLTGLACGVASLVAPVYISEIAYPAVRGLLGSCVQLMVVTGILLAYVAGWVLE 180

Qy         181 WRWLAVLGCVPPSLMLLLMCFMPETPRFLLTQHRRQEAMAALRFLWGSEQGWEDPPIGAE 240
               ||||||||||||:|||||||:||||||||||||: ||||||||||||||:|||:||:|||
Db         181 WRWLAVLGCVPPTLMLLLMCYMPETPRFLLTQHQYQEAMAALRFLWGSEEGWEEPPVGAE 240

Qy         241 -QSFHLALLRQPGIYKPFIIGVSLMAFQQLSGVNAVMFYAETIFEEAKFKDSSLASVVVG 299
                | | ||:||:||::|| |||: || |||||||||:|||| |||||||||||||||| ||
Db         241 HQGFQLAMLRRPGVHKPLIIGICLMVFQQLSGVNAIMFYANTIFEEAKFKDSSLASVTVG 300

Qy         300 VIQVLFTAVAALIMDRAGRRLLLVLSGVVMVFSTSAFGAYFKLTQGGPGNSSHVAISAPV 359
               :||||||||||||||||||:||| ||||:|||| |||| |||||| || ||||| :  |:
Db         301 IIQVLFTAVAALIMDRAGRKLLLALSGVIMVFSMSAFGTYFKLTQSGPSNSSHVGLLVPI 360

Qy         360 SAQPVDASVGLAWLAVGSMCLFIAGFAVGWGPIPWLLMSEIFPLHVKGVATGICVLTNWL 419
               ||:| |  :||||||||||||||||||||||||||||||||||||:||||||:||||||
Db         361 SAEPADVHLGLAWLAVGSMCLFIAGFAVGWGPIPWLLMSEIFPLHIKGVATGVCVLTNWF 420

Qy         420 MAFLVTKEFSSLMEVLRPYGAFWLASAFCIFSVLFTLFCVPETKGKTLEQITAHFEGR 477
               |||||||||:|:||:||||||||| :|||| ||||||  ||||||:||||||||||||
Db         421 MAFLVTKEFNSIMEILRPYGAFWLTAAFCILSVLFTLTFVPETKGRTLEQITAHFEGR 478
```

```
RESULT 6
AAB66934
ID   AAB66934 standard; protein; 477 AA.
XX
AC   AAB66934;
XX
DT   17-APR-2001  (first entry)
XX
DE   Murine GLUTX1.
XX
KW   Murine; GLUTX; gene therapy; vaccine; hexose transport modulator;
KW   hexose transport disorder; ischaemia; diabetes; hyperglycaemia;
KW   hypoglycaemia; glucose metabolism disorder; neurodegenerative disease.
XX
OS   Mus sp.
XX
PN   WO200104145-A2.
XX
PD   18-JAN-2001.
XX
PF   14-JUL-2000; 2000WO-IB001042.
XX
PR   14-JUL-1999;   99US-0143907P.
PR   27-AUG-1999;   99US-0151140P.
PR   23-FEB-2000; 2000US-0184285P.
PR   13-JUL-2000; 2000US-00616132.
XX
PA   (UYLA-) UNIV LAUSANNE.
XX
PI   Thorens B,  Ibberson M,  Uldry M;
XX
DR   WPI; 2001-112615/12.
DR   N-PSDB; AAF55867.
XX
PT   Nucleic acids encoding GLUTX glucose transporter proteins, useful in the
PT   prevention, diagnosis and treatment of hexose transport disorders, e.g.
PT   ischemia and diabetes.
XX
PS   Claim 11; Page 73-74; 124pp; English.
XX
CC   The present invention relates to GLUTX proteins (AAF55865-AAF55871 and
CC   AAB66932-AAB66941). The GLUTX proteins are related to the facultative
CC   glucose carriers GLUT1-GLUT5 and have hexose binding and/or transport
CC   function. The GLUTX proteins may be used in the diagnosis, prevention and
CC   treatment of hexose transport disorders such as ischaemia, diabetes,
CC   hyperglycaemia, hypoglycaemia, a glucose metabolism disorder and/or a
CC   neurodegenerative disease. The present sequence is murine GLUTX1
XX
SQ   Sequence 477 AA;

  Query Match             87.5%;  Score 2150;  DB 4;  Length 477;
  Best Local Similarity   85.6%;  Pred. No. 1.6e-197;
  Matches  409;  Conservative   33;  Mismatches   34;  Indels    2;  Gaps    2;

Qy           1 MTPEDPEETQPLLGPPGGSAPRGRRVFLAAFAAALGPLSFGFALGYSSPAIPSLQRAAPP 60
               |:||||:|||||| ||    |||||||||:||||||||:|||||||||||||||:| |||
Db           1 MSPEDPQETQPLLRPPEARTPRGRRVFLASFAAALGPLNFGFALGYSSPAIPSLRRTAPP 60

Qy          61 APRLDDAAASWFGAVVTLGAAAGGVLGGWLVDRAGRKLSLLLCSVPFVAGFAVITAAQDV 120
               | || | |||||||||||||||||:|||||:||:|||||||||:|||| ||||||||:||
Db          61 ALRLGDNAASWFGAVVTLGAAAGGILGGWLLDRSGRKLSLLLCTVPFVTGFAVITAARDV 120

Qy         121 WMLLGGRLLTGLACGVASLVAPVYISEIAYPAVRGLLGSCVQLMVVVGILLAYLAGWVLE 180
               |||||||||||||||||||||||||||||||||||||||||||||| ||||||:||||||
Db         121 WMLLGGRLLTGLACGVASLVAPVYISEIAYPAVRGLLGSCVQLMVVTGILLAYVAGWVLE 180

Qy         181 WRWLAVLGCVPPSLMLLLMCFMPETPRFLLTQHRRQEAMAALRFLWGSEQGWEDPPIGAE 240
               ||||||||||||:|||||||:||||||||||||: ||||||||||||||:|||:||:|||
Db         181 WRWLAVLGCVPPTLMLLLMCYMPETPRFLLTQHQYQEAMAALRFLWGSEEGWEEPPVGAE 240

Qy         241 -QSFHLALLRQPGIYKPFIIGVSLMAFQQLSGVNAVMFYAETIFEEAKFKDSSLASVVVG 299
                | | |||||:|||||| |||:||| |||||||||:|||| :||||||||||||||| ||
Db         241 HQGFQLALLRRPGIYKPLIIGISLMVFQQLSGVNAIMFYANSIFEEAKFKDSSLASVTVG 300

Qy         300 VIQVLFTAVAALIMDRAGRRLLLVLSGVVMVFSTSAFGAYFKLTQGGPGNSSHVAISAPV 359
               :|||||||||||||||||||||| ||||:|||| |||| ||||||  | ||||| :  |:
Db         301 IIQVLFTAVAALIMDRAGRRLLLALSGVIMVFSMSAFGTYFKLTQSLPSNSSHVGL-VPI 359

Qy         360 SAQPVDASVGLAWLAVGSMCLFIAGFAVGWGPIPWLLMSEIFPLHVKGVATGICVLTNWL 419
               :|:|||  |||||||||||||||||||||||||||||||||||||||||||||||||||
Db         360 AAEPVDVQVGLAWLAVGSMCLFIAGFAVGWGPIPWLLMSEIFPLHVKGVATGICVLTNWF 419

Qy         420 MAFLVTKEFSSLMEVLRPYGAFWLASAFCIFSVLFTLFCVPETKGKTLEQITAHFEGR 477
               |||||||||:|:||:||||||||| :|||  ||||||  ||||||:||||:|||||||
Db         420 MAFLVTKEFNSVMEMLRPYGAFWLTAAFCALSVLFTLTVVPETKGRTLEQVTAHFEGR 477
```

```
RESULT 7
ABB89717
ID   ABB89717 standard; protein; 326 AA.
XX
AC   ABB89717;
XX
DT   24-MAY-2002  (first entry)
XX
DE   Human polypeptide SEQ ID NO 2093.
XX
KW   Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW   antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
KW   vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW   cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW   neurological disease; infection; human; secreted protein.
XX
OS   Homo sapiens.
XX
PN   WO200190304-A2.
XX
PD   29-NOV-2001.
XX
PF   18-MAY-2001; 2001WO-US016450.
XX
PR   19-MAY-2000; 2000US-0205515P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Birse CE,  Rosen CA;
```

```
XX
DR   WPI; 2002-122018/16.
DR   N-PSDB; ABL90126.
XX
PT   Novel 1405 isolated polypeptides, useful for diagnosis, treatment and
PT   prevention of neural, immune system, muscular, reproductive,
PT   gastrointestinal, pulmonary, cardiovascular, renal and proliferative
PT   disorders.
XX
PS   Claim 11; SEQ ID NO 2093; 2081pp + Sequence Listing; English.
XX
CC   The invention relates to novel genes (ABL89449-ABL90853) and proteins
CC   (ABB89040-ABB90444) useful for preventing, treating or ameliorating
CC   medical conditions e.g. by protein or gene therapy. The genes are
CC   isolated from a range of human tissues disclosed in the specification.
CC   The nucleic acids, proteins, antibodies and (ant)agonists are useful in
CC   the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
CC   ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,
CC   breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune
CC   disorders e.g. Addison's disease, allergies, autoimmune haemolytic
CC   anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
CC   multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
CC   cardiovascular disorders such as myocardial ischaemias; (d) wound healing
CC   ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
CC   infectious diseases such as viral, bacterial, fungal and parasitic
CC   infections. Note: The sequence data for this patent did not form part of
CC   the printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 326 AA;

  Query Match             60.7%;  Score 1491;  DB 5;  Length 326;
  Best Local Similarity   96.3%;  Pred. No. 1.9e-134;
  Matches  289;  Conservative    1;  Mismatches   10;  Indels    0;  Gaps    0;

Qy        164 MVVVGILLAYLAGWVLEWRWLAVLGCVPPSLMLLLMCFMPETPRFLLTQHRRQEAMAALR 223
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MVVVGILLAYLAGWVLEWRWLAVLGCVPPSLMLLLMCFMPETPRFLLTQHRRQEAMAALR 60

Qy        224 FLWGSEQGWEDPPIGAEQSFHLALLRQPGIYKPFIIGVSLMAFQQLSGVNAVMFYAETIF 283
              ||||||||||||||| |||||||||| |||||||||||||||||||||||||||||||||
Db         61 FLWGSEQGWEDPPIGXEQSFHLALLRXPGIYKPFIIGVSLMAFQQLSGVNAVMFYAETIF 120

Qy        284 EEAKFKDSSLASVVVGVIQVLFTAVAALIMDRAGRRLLLVLSGVVMVFSTSAFGAYFKLT 343
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 EEAKFKDSSLASVVVGVIQVLFTAVAALIMDRAGRRLLLVLSGVVMVFSTSAFGAYFKLT 180

Qy        344 QGGPGNSSHVAISAPVSAQPVDASVGLAWLAVGSMCLFIAGFAVGWGPIPWLLMSEIFPL 403
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 QGGPGNSSHVAISAPVSAQPVDASVGLAWLAVGSMCLFIAGFAVGWGPIPWLLMSEIFPL 240

Qy        404 HVKGVATGICVLTNWLMAFLVTKEFSSLMEVLRPYGAFWLASAFCIFSVLFTLFCVPETK 463
              |||||||| ||| |||| || |||||||||||||||||||||||||||||||| |  : |
Db        241 HVKGVATGXCVLXNWLMXFLXTKEFSSLMEVLRPYGAFWLASAFCIFSVLFTLVCALKLK 300


RESULT 8
AAE06579
ID   AAE06579 standard; protein; 262 AA.
XX
AC   AAE06579;
XX
DT   25-SEP-2001  (first entry)
XX
DE   Human protein having hydrophobic domain, HP10784.
XX
KW   Human; hydrophobic domain; gene therapy; nutritional supplement;
KW   cell proliferation; immunomodulatory; autoimmune disorder; antimicrobial;
KW   multiple sclerosis; rheumatoid arthritis; insulin-dependent diabetes;
KW   haematopoiesis; tissue growth activity; Parkinson's disease; cytostatic;
KW   Huntington's disease; Alzheimer's disease; chemotactic; chemokinetic;
KW   haemostatic; thrombolytic; tumour growth inhibitor; anabolic;
KW   contraceptive; antiinfertility; antiinflammatory.
XX
OS   Homo sapiens.
XX
PN   WO200149728-A2.
XX
PD   12-JUL-2001.
XX
PF   28-DEC-2000; 2000WO-JP009359.
XX
PR   06-JAN-2000; 2000JP-00000585.
PR   06-JAN-2000; 2000JP-00000588.
PR   11-JAN-2000; 2000JP-00002299.
PR   03-FEB-2000; 2000JP-00026862.
PR   03-MAR-2000; 2000JP-00058367.
XX
PA   (PROT-) PROTEGENE INC.
PA   (SAGA ) SAGAMI CHEM RES CENT.
XX
PI   Kato S,  Kimura T;
XX
DR   WPI; 2001-418355/44.
DR   N-PSDB; AAD12574.
XX
PT   Human proteins with hydrophobic domains and the nucleic acids encoding
PT   them, useful for preventing diagnosing and treating e.g. cancer,
PT   Alzheimer's and inflammation.
XX
PS   Claim 1; Page 75; 563pp; English.
XX
CC   The present sequence is human protein with hydrophobic domain, HP10784.
CC   The polynucleotide and polypeptide of the invention may be used in the
CC   prevention, diagnosis and treatment of diseases associated with
CC   inappropriate polypeptide expression. The polynucleotides may be used to
CC   produce the polypeptide, by inserting the nucleic acids into a host cell
CC   and culturing the cell to express the protein. The polynucleotides and
CC   its complementary sequences may also be used as DNA probes in diagnostic
CC   assays and also used in gene therapy. The polypeptides may also be used
CC   as antigens in the production of antibodies and in assays to identify
CC   modulators of polypeptide expression and activity. The polypeptides and
CC   nucleic acids may be used as nutritional supplements, to modulate
CC   cytokine and cell proliferation activity, to modulate immune stimulation
CC   or suppression (e.g. for the treatment of microbial infections and
CC   autoimmune disorders such as multiple sclerosis, rheumatoid arthritis and
CC   insulin-dependent diabetes), to modulate haematopoiesis, to modulate
CC   tissue growth activity (e.g. for the treatment of Parkinson's disease,
CC   Huntington's disease and Alzheimer's disease), to modulate activin and
CC   inhibin activity (e.g. for controlling fertility), to modulate
CC   chemotactic and chemokinetic activity, to modulate haemostatic and
CC   thrombolytic activity, to modulate receptor ligand activity, to modulate
CC   inflammation and to inhibit tumour growth
XX
SQ   Sequence 262 AA;

  Query Match             46.2%;  Score 1135;  DB 4;  Length 262;
  Best Local Similarity   91.4%;  Pred. No. 2.2e-100;
  Matches  223;  Conservative    2;  Mismatches    7;  Indels   12;  Gaps    1;

Qy          1 MTPEDPEETQPLLGPPGGSAPRGRRVFLAAFAAALGPLSFGFALGYSSPAIPSLQRAAPP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MTPEDPEETQPLLGPPGGSAPRGRRVFLAAFAAALGPLSFGFALGYSSPAIPSLQRAAPP 60

Qy         61 APRLDDAAASWFGAVVTLGAAAGGVLGGWLVDRAGRKLSLLLCSVPFVAGFAVITAAQDV 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 APRLDDAAASWFGAVVTLGAAAGGVLGGWLVDRAGRKLSLLLCSVPFVAGFAVITAAQDV 120

Qy        121 WMLLGGRLLTGLACGVASLVAPVYISEIAYPAVRGLLGSCVQLMVVVGILLAYLAGWVLE 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 WMLLGGRLLTGLACGVASLVAPVYISEIAYPAVRGLLGSCVQLMVVVGILLAYLAGWVLE 180

Qy        181 WRWLAVLGCVPPSLMLLLMCFMPETPRFLLTQHRRQEAM------------AALRFLWGS 228
              ||||||||||||||||||||||||||||||||||||||               || |  :
Db        181 WRWLAVLGCVPPSLMLLLMCFMPETPRFLLTQHRRQEAAPGLVRCGHGVQHECLRRLLQA 240
```

```
Qy        229 EQGW 232
              : ||
Db        241 DPGW 244

RESULT 9
ABU11283
ID   ABU11283 standard; protein; 248 AA.
XX
AC   ABU11283;
XX
DT   10-FEB-2003  (first entry)
XX
DE   cDNA encoding human cancer suppressing protein PP7425.
XX
KW   Human; cancer suppressing protein; cancer.
XX
OS   Homo sapiens.
XX
PN   CN1351081-A.
XX
PD   29-MAY-2002.
XX
PF   31-OCT-2000; 2000CN-00127102.
XX
PR   31-OCT-2000; 2000CN-00127102.
XX
PA   (SHAN-) SHANGHAI INST ONCOLOGY.
XX
PI   Gu J;
XX
DR   WPI; 2002-609437/66.
DR   N-PSDB; ABX34029.
XX
PT   New human protein with cancer cell growth suppressing function and a
PT   polynucleotide encoding it, for treating diseases, such as, cancer.
XX
PS   Claim 1; Page 23 (disclosure); 39pp; Chinese.
XX
CC   This invention relates to the cDNA and protein sequences of a novel human
CC   protein with cancer suppressing function. The invention also comprises a
CC   method for preparing the polypeptide by recombination, and an application
CC   of the polypeptide in treating diseases such as cancer, etc. Also
CC   disclosed in an antagonist of the polypeptide and its medical action. The
CC   present sequence represents a cancer suppressing protein of the invention
XX
SQ   Sequence 248 AA;

  Query Match             46.0%;  Score 1131;  DB 5;  Length 248;
  Best Local Similarity   93.0%;  Pred. No. 5e-100;
  Matches  227;  Conservative    5;  Mismatches    6;  Indels    6;  Gaps    2;

Qy        164 MVVVGILLAYLAGWVLEWRWLAVLGCVPPSLMLLLMCFMPETPRFLLTQHRRQEAMAALR 223
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MVVVGILLAYLAGWVLEWRWLAVLGCVPPSLMLLLMCFMPETPRFLLTQHRRQEAMAALR 60

Qy        224 FLWGSEQGWEDPPIGAEQSFHLALLRQPGIYKPFIIGVSLMAFQQLSGVNAVMFYAETIF 283
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 FLWGSEQGWEDPPIGAEQSFHLALLRQPGIYKPFIIGVSLMAFQQLSGVNAVMFYAETIF 120

Qy        284 EEAKFKDSSLASVVVGVIQVLFTAVAALIMDRAGRRLLLVLSGVVMVFSTSAFGAYFKLT 343
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 EEAKFKDSSLASVVVGVIQVLFTAVAALIMDRAGRRLLLVLSGVVMVFSTSAFGAYFKLT 180

Qy        344 QGGPGNSSHVAISAPVSAQPVDASVGLAWLAVGSMCLFIAGFAVGWGP-IPWLLMSEIFP 402
              |||||||||||||||||||||||||||||||||||||||||     ||   | |:: :
Db        181 QGGPGNSSHVAISAPVSAQPVDASVGLAWLAVGSMCLFIAG-----GPQALWSLLACLRF 235

Qy        403 LHVK 406
              ||::
Db        236 LHLQ 239

RESULT 10
AAB66938
ID   AAB66938 standard; protein; 503 AA.
XX
AC   AAB66938;
XX
DT   17-APR-2001  (first entry)
XX
DE   Rat GLUTX3.
XX
KW   Rat; GLUTX; gene therapy; vaccine; hexose transport modulator;
KW   hexose transport disorder; ischaemia; diabetes; hyperglycaemia;
KW   hypoglycaemia; glucose metabolism disorder; neurodegenerative disease.
XX
OS   Rattus sp.
XX
PN   WO200104145-A2.
XX
PD   18-JAN-2001.
XX
PF   14-JUL-2000; 2000WO-IB001042.
XX
PR   14-JUL-1999;   99US-0143907P.
PR   27-AUG-1999;   99US-0151140P.
PR   23-FEB-2000; 2000US-0184285P.
PR   13-JUL-2000; 2000US-00616132.
XX
PA   (UYLA-) UNIV LAUSANNE.
XX
PI   Thorens B,  Ibberson M,  Uldry M;
XX
DR   WPI; 2001-112615/12.
DR   N-PSDB; AAF55871.
XX
PT   Nucleic acids encoding GLUTX glucose transporter proteins, useful in the
PT   prevention, diagnosis and treatment of hexose transport disorders, e.g.
PT   ischemia and diabetes.
XX
PS   Claim 11; Page 82-83; 124pp; English.
XX
CC   The present invention relates to GLUTX proteins (AAF55865-AAF55871 and
CC   AAB66932-AAB66941). The GLUTX proteins are related to the facultative
CC   glucose carriers GLUT1-GLUT5 and have hexose binding and/or transport
CC   function. The GLUTX proteins may be used in the diagnosis, prevention and
CC   treatment of hexose transport disorders such as ischaemia, diabetes,
CC   hyperglycaemia, hypoglycaemia, a glucose metabolism disorder and/or a
CC   neurodegenerative disease. The present sequence is rat GLUTX3
XX
SQ   Sequence 503 AA;

  Query Match             39.5%;  Score 969.5;  DB 4;  Length 503;
  Best Local Similarity   43.6%;  Pred. No. 4.2e-84;
  Matches  212;  Conservative   80;  Mismatches  175;  Indels   19;  Gaps    7;

Qy          3 PEDPEETQPLLGPPGGSAPRGRRVFLAAFAAALGPLSFGFALGYSSPAIPSLQRAAPPAP 62
              |  |||           | : |||||| ||| ||  |||:|| |:|| ||:|:|:: ||
Db         20 PASPEE------KARAGALQNRRVFLATFAAVLGNFSFGYALVYTSPVIPALKRSSDPAL 73

Qy         63 RLDDAAASWFGAVVTLGAAAGGVLGGWLVDRAGRKLSLLLCSVPFVAGFAVITAAQDVWM 122
              |||   |||||:| ||||||||:    | |  |||||::  :||   |:|::  |: :||
Db         74 RLDKIQASWFGSVFTLGAAAGGLSAMLLNDLLGRKLSIMFSAVPSAIGYALMAGARGLWM 133

Qy        123 LLGGRLLTGLACGVASLVAPVYISEIAYPAVRGLLGSCVQLMVVVGILLAYLAGWVLEWR 182
              || ||:||| | |: :   |||:|||| | ||| ||:  ||| | | |  |  | :| ||
Db        134 LLLGRMLTGFAGGLTAACIPVYVSEIAPPGVRGALGATPQLMAVFGSLSLYALGLLLPWR 193

Qy        183 WLAVLGCVPPSLMLLLMCFMPETPRFLLTQHRRQEAMAALRFLW---GSEQGWEDPPIG- 238
              |||| |  |  :|:||: ||| :|||||:: | :||: ||  :|    ||  ||   |
Db        194 WLAVAGEGPVLVMILLLSFMPNSPRFLLSKSRDEEALQAL--IWLRADSEVHWEFEQIQD 251

Qy        239 --AEQSFHLALLR--QPGIYKPFIIGVSLMAFQQLSGVNAVMFYAETIFEEAK-FKDSSL 293
                  ||  ::     :| :|:| :| | :   |||:|:  :: | :|||:       |
Db        252 NVRRQSSRVSWAEAWEPRVYRPILITVLMRFLQQLTGITPILVYLQTIFDSTSVVLPSQQ 311
```

```
Qy          294 ASVVVGVIQVLFTAVAALIMDRAGRRLLLVLSGVVMVFSTSAFGAYFKLTQG--GPGNSS 351
                 : :|| :::|   :||: || |||::|| :|  :|  :    | | :|      | ::
Db          312 DAAIVGAVRLLSVLIAAVTMDLAGRKVLLYVSASIMFVANLTLGLYVQLVPRTLTPNSTV 371

Qy          352 HVAISAPVSAQPVDASVGLAWLAVGSMCLFIAGFAVGWGPIPWLLMSEIFPLHVKGVATG 411
                 :         |  |   |  : : :  ||| |:|:||||| ||||||: ||  :|||:|
Db          372 EIVTLGGTEQPPAAAFNYLTLIPLLATMLFIMGYAMGWGPITWLLMSEVLPLRARGVASG 431

Qy          412 ICVLTNWLMAFLVTKEFSSLMEVLRPYGAFWLASAFCIFSVLFTLFCVPETKGKTLEQIT 471
                :||| :|| ||::|| |   :        |:  || |: |:|||  |||||:|::||||
Db          432 LCVLVSWLTAFVLTKYFLLAVNAFGLQVPFFFFSAICLLSLLFTGCCVPETRGRSLEQIE 491

Qy          472 AHFEGR 477
                | |  |
Db          492 AFFHTR 497

RESULT 11
AAB66941
ID   AAB66941 standard; protein; 507 AA.
XX
AC   AAB66941;
XX
DT   17-APR-2001  (first entry)
XX
DE   GLUTX3 consensus sequence.
XX
KW   GLUTX; gene therapy; vaccine; hexose transport modulator; human; rat;
KW   hexose transport disorder; ischaemia; diabetes; hyperglycaemia;
KW   hypoglycaemia; glucose metabolism disorder; neurodegenerative disease.
XX
OS   Homo sapiens.
OS   Rattus sp.
XX
PN   WO200104145-A2.
XX
PD   18-JAN-2001.
XX
PF   14-JUL-2000; 2000WO-IB001042.
XX
PR   14-JUL-1999;   99US-0143907P.
PR   27-AUG-1999;   99US-0151140P.
PR   23-FEB-2000; 2000US-0184285P.
PR   13-JUL-2000; 2000US-00616132.
XX
PA   (UYLA-) UNIV LAUSANNE.
XX
PI   Thorens B,  Ibberson M,  Uldry M;
XX
DR   WPI; 2001-112615/12.
XX
PT   Nucleic acids encoding GLUTX glucose transporter proteins, useful in the
PT   prevention, diagnosis and treatment of hexose transport disorders, e.g.
PT   ischemia and diabetes.
XX
PS   Claim 11; Page 83-84; 124pp; English.
XX
CC   The present invention relates to GLUTX proteins (AAF55865-AAF55871 and
CC   AAB66932-AAB66941). The GLUTX proteins are related to the facultative
CC   glucose carriers GLUT1-GLUT5 and have hexose binding and/or transport
CC   function. The GLUTX proteins may be used in the diagnosis, prevention and
CC   treatment of hexose transport disorders such as ischaemia, diabetes,
CC   hyperglycaemia, hypoglycaemia, a glucose metabolism disorder and/or a
CC   neurodegenerative disease. The present sequence is a consensus sequence
CC   for GLUTX3
XX
SQ   Sequence 507 AA;

  Query Match             38.8%;  Score 953;  DB 4;  Length 507;
  Best Local Similarity   42.9%;  Pred. No. 1.6e-82;
  Matches  217;  Conservative   78;  Mismatches  167;  Indels   44;  Gaps   11;

Qy           10 QPLLG------------PPGGSAPRGR-------RVFLAAFAAALGPLSFGFALGYSSPA 50
                :||||            ||     | |       ||||| ||| ||  |||:|| |:||
Db            3 EPLLGAEGPDYDTFPEKPPASPGDRARVGTLQNKRVFLATFAAVLGNFSFGYALVYTSPV 62

Qy           51 IPSLQRAAPPAPRLDDAAASWFGAVVTLGAAAGGVLGGWLVDRAGRKLSLLLCSVPFVAG 110
                ||:|:|:  |   |  : |||||:| ||||||||:    | |  |||||::  :||  ||
Db           63 IPALERSLDPDLHLTKSQASWFGSVFTLGAAAGGLSAMILNDLLGRKLSIMFSAVPSAAG 122

Qy          111 FAVITAAQDVWMLLGGRLLTGLACGVASLVAPVYISEIAYPAVRGLLGSCVQLMVVVGIL 170
                :||:  |  :|||| || ||| | |: :   |||:|||| | ||| ||:  ||| | | |
Db          123 YAVMAGAHGLWMLLLGRTLTGFAGGLTAACIPVYVSEIAPPGVRGALGATPQLMAVFGSL 182

Qy          171 LAYLAGWVLEWRWLAVLGCVPPSLMLLLMCFMPETPRFLLTQHRRQEAMAALRFLWGSEQ 230
                  |  | :| |||||| |  |  :|:||: ||| :|||||:: | :||: || :| |::
Db          183 SLYALGLLLPWRWLAVAGXAPVLIMILLLSFMPNSPRFLLSRGRDEEALRALAWLRGTDV 242

Qy          231 G--WEDPPIG---AEQSFHL--ALLRQPGIYKPFIIGVSLMAFQQLSGVNAVMFYAETIF 283
                   ||   |      ||  :  |  | | : :|  : : :   |||:|:  :: | ::||
Db          243 DVHWEFGQIQDNVRRQSSRVSWAEARAPHVCRPITVALLMRLLQQLTGITPILVYLQSIF 302

Qy          284 EEAKF----KDSSLASVVVGVIQVLFTAVAALIMDRAGRRLLLVLSGVVMVFSTSAFGAY 339
                :        ||    : :|| :::|   :||| || |||::|| :|  :|  :    | |
Db          303 DSTAVLLPPKDD---AAIVGAVRLLSVLIAALTMDLAGRKVLLFVSAAIMFAANLTLGLY 359

Qy          340 FKLTQGGPGNSSHVAISAPVS------AQPVDASVG-LAWLAVGSMCLFIAGFAVGWGPI 392
                      ||   |  : :   |      |||: |  | |  : : :  ||| |:|||||||
Db          360 IHF---GPRRLSPNSTAGLESESWGDLAQPLAAPAGYLTLVPLLATMLFIMGYAVGWGPI 416

Qy          393 PWLLMSEIFPLHVKGVATGICVLTNWLMAFLVTKEFSSLMEVLRPYGAFWLASAFCIFSV 452
                 ||||||: ||  :|||:|:||| :|| ||::|| |  ::        |   :| |: |:
Db          417 TWLLMSEVLPLRARGVASGLCVLASWLTAFVLTKSFLPVVSTFGLQVPFLFFAAICLVSL 476

Qy          453 LFTLFCVPETKGKTLEQITAHFE-GR 477
                :||  |||||||::|||| : |  ||
Db          477 VFTGCCVPETKGRSLEQIESFFRTGR 502

RESULT 12
AAB66937
ID   AAB66937 standard; protein; 507 AA.
XX
AC   AAB66937;
XX
DT   17-APR-2001  (first entry)
XX
DE   Human GLUTX3.
XX
KW   Human; GLUTX; gene therapy; vaccine; hexose transport modulator;
KW   hexose transport disorder; ischaemia; diabetes; hyperglycaemia;
KW   hypoglycaemia; glucose metabolism disorder; neurodegenerative disease.
XX
OS   Homo sapiens.
XX
PN   WO200104145-A2.
XX
PD   18-JAN-2001.
XX
PF   14-JUL-2000; 2000WO-IB001042.
XX
PR   14-JUL-1999;   99US-0143907P.
PR   27-AUG-1999;   99US-0151140P.
PR   23-FEB-2000; 2000US-0184285P.
PR   13-JUL-2000; 2000US-00616132.
XX
PA   (UYLA-) UNIV LAUSANNE.
XX
PI   Thorens B,  Ibberson M,  Uldry M;
XX
DR   WPI; 2001-112615/12.
DR   N-PSDB; AAF55870.
XX
PT   Nucleic acids encoding GLUTX glucose transporter proteins, useful in the
PT   prevention, diagnosis and treatment of hexose transport disorders, e.g.
```

```
PT   ischemia and diabetes.
XX
PS   Claim 11; Page 81-82; 124pp; English.
XX
CC   The present invention relates to GLUTX proteins (AAF55865-AAF55871 and
CC   AAB66932-AAB66941). The GLUTX proteins are related to the facultative
CC   glucose carriers GLUT1-GLUT5 and have hexose binding and/or transport
CC   function. The GLUTX proteins may be used in the diagnosis, prevention and
CC   treatment of hexose transport disorders such as ischaemia, diabetes,
CC   hyperglycaemia, hypoglycaemia, a glucose metabolism disorder and/or a
CC   neurodegenerative disease. The present sequence is human GLUTX3
XX
SQ   Sequence 507 AA;

  Query Match             38.6%;  Score 948;  DB 4;  Length 507;
  Best Local Similarity   42.9%;  Pred. No. 4.9e-82;
  Matches  217;  Conservative   78;  Mismatches  167;  Indels   44;  Gaps   11;

Qy         10 QPLLG------------PPGGSAPRGR-------RVFLAAFAAALGPLSFGFALGYSSPA 50
              :||||            ||     | |       ||||| ||| ||  |||:|| |:||
Db          3 EPLLGAEGPDYDTFPEKPPPSPGDRARVGTLQNKRVFLATFAAVLGNFSFGYALVYTSPV 62

Qy         51 IPSLQRAAPPAPRLDDAAASWFGAVVTLGAAAGGVLGGWLVDRAGRKLSLLLCSVPFVAG 110
              ||:|:|:  |   |  : |||||:| ||||||||:    | |  |||||::  :||  ||
Db         63 IPALERSLDPDLHLTKSQASWFGSVFTLGAAAGGLSAMILNDLLGRKLSIMFSAVPSAAG 122

Qy        111 FAVITAAQDVWMLLGGRLLTGLACGVASLVAPVYISEIAYPAVRGLLGSCVQLMVVVGIL 170
              :||:  |  :|||| || ||| | |: :   |||:|||| | ||| ||:  ||| | | |
Db        123 YAVMAGAHGLWMLLLGRTLTGFAGGLTAACIPVYVSEIAPPGVRGALGATPQLMAVFGSL 182

Qy        171 LAYLAGWVLEWRWLAVLGCVPPSLMLLLMCFMPETPRFLLTQHRRQEAMAALRFLWGSEQ 230
                |  | :| |||||| |  |  :|:||: ||| :|||||:: | :||: || :| |::
Db        183 SLYALGLLLPWRWLAVAGXAPVLIMILLLSFMPNSPRFLLSRGRDEEALRALAWLRGTDV 242

Qy        231 G--WEDPPIG---AEQSFHL--ALLRQPGIYKPFIIGVSLMAFQQLSGVNAVMFYAETIF 283
                 ||   |      ||  :  |  | | : :|  : : :   |||:|:  :: | ::||
Db        243 DVHWEFGQIQDNVRRQSSRVSWAEARAPHVCRPITVALLMRLLQQLTGITPILVYLQSIF 302

Qy        284 EEAKF----KDSSLASVVVGVIQVLFTAVAALIMDRAGRRLLLVLSGVVMVFSTSAFGAY 339
              :        ||    : :|| :::|   :||| || |||::|| :|  :|   :    | |
Db        303 DSTAVLLPPKDD---AAIVGXVRLLSVLIAALTMDLAGRKVLLFVSAAIMFAANLTLGLY 359

Qy        340 FKLTQGGPGNSSHVAISAPVS------AQPVDASVG-LAWLAVGSMCLFIAGFAVGWGPI 392
                    ||   |  : :   |      |||: |  | |  : : :  ||| |:|||||||
Db        360 IHF---GPRPLSPNSTAGLESESWGDLAQPLAAPAGYLTLVPLLATMLFIMGYAVGWGPI 416

Qy        393 PWLLMSEIFPLHVKGVATGICVLTNWLMAFLVTKEFSSLMEVLRPYGAFWLASAFCIFSV 452
               ||||||: ||  :|||:|:||| :|| ||::|| |  ::        |   :| |: |:
Db        417 TWLLMSEVLPLRARGVASGLCVLASWLTAFVLTKSFLPVVSTFGLQVPFLFFAAICLVSL 476

Qy        453 LFTLFCVPETKGKTLEQITAHFE-GR 477
              :||  |||||||::|||| : |  ||
Db        477 VFTGCCVPETKGRSLEQIESFFRTGR 502


RESULT 13
AAM93417
ID   AAM93417 standard; protein; 507 AA.
XX
AC   AAM93417;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human polypeptide, SEQ ID NO: 3034.
XX
KW   Human; full length cDNA; cDNA synthesis; oligo-capping.
XX
OS   Homo sapiens.
XX
PN   EP1130094-A2.
XX
PD   05-SEP-2001.
XX
```

```
PF   07-JUL-2000; 2000EP-00114089.
XX
PR   08-JUL-1999;   99JP-00194486.
PR   11-JAN-2000; 2000JP-00118774.
PR   02-MAY-2000; 2000JP-00183765.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Nishikawa T,  Isogai T,  Hayashi K,  Ishii S,  Kawai Y;
PI   Wakamatsu A,  Sugiyama T,  Nagai K,  Kojima S,  Otsuki T,  Koga H;
XX
DR   WPI; 2001-524255/58.
DR   N-PSDB; AAK94338.
XX
PT   830 Primers useful for synthesizing full length cDNA clones and their use
PT   in genetic manipulation.
XX
PS   Claim 8; SEQ ID NO 3034; 1380pp + Sequence Listing; English.
XX
CC   The invention relates to primers for synthesising full length cDNA
CC   clones. 830 cDNA molecules encoding a human protein have been isolated
CC   and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have
CC   been determined. Primers for synthesising the full length cDNA are useful
CC   for clarifying the function of the protein encoded by the cDNA. The full
CC   length clones were obtained by construction of full length enriched cDNA
CC   libraries that were synthesised by the oligo-capping method. The primers
CC   enable the production of the full length cDNA easily without any special
CC   methods. The present sequence is a polypeptide encoded by a full length
CC   human cDNA of the invention. Note: The sequence data for this patent did
CC   not form part of the printed specification, but was obtained in CD-ROM
CC   format directly from EPO
XX
SQ   Sequence 507 AA;

  Query Match             38.6%;  Score 948;  DB 4;  Length 507;
  Best Local Similarity   42.7%;  Pred. No. 4.9e-82;
  Matches  216;  Conservative   80;  Mismatches  166;  Indels   44;  Gaps   11;

Qy         10 QPLLG------------PPGGSAPRGR-------RVFLAAFAAALGPLSFGFALGYSSPA 50
              :||||            ||     | |       ||||| ||| ||  |||:|| |:||
Db          3 EPLLGAEGPDYDTFPEKPPPSPGDRARVGTLQNKRVFLATFAAVLGNFSFGYALVYTSPV 62

Qy         51 IPSLQRAAPPAPRLDDAAASWFGAVVTLGAAAGGVLGGWLVDRAGRKLSLLLCSVPFVAG 110
              ||:|:|:  |   |  : |||||:| ||||||||:    | |  |||||::  :||  ||
Db         63 IPALERSLDPDLHLTKSQASWFGSVFTLGAAAGGLSAMILNDLLGRKLSIMFSAVPSAAG 122

Qy        111 FAVITAAQDVWMLLGGRLLTGLACGVASLVAPVYISEIAYPAVRGLLGSCVQLMVVVGIL 170
              :|::  |  :|||| || ||| | |: :   |||:|||| | ||| ||:  ||| | | |
Db        123 YALMAGAHGLWMLLLGRTLTGFAGGLTAACIPVYVSEIAPPGVRGALGATPQLMAVFGSL 182

Qy        171 LAYLAGWVLEWRWLAVLGCVPPSLMLLLMCFMPETPRFLLTQHRRQEAMAALRFLWGSEQ 230
                |  | :| |||||| |  |  :|:||: ||| :|||||:: | :||: || :| |::
Db        183 SLYALGLLLPWRWLAVAGEAPVLIMILLLSFMPNSPRFLLSRGRDEEALRALAWLRGTDV 242

Qy        231 G--WEDPPIG---AEQSFHL--ALLRQPGIYKPFIIGVSLMAFQQLSGVNAVMFYAETIF 283
                 ||   |      ||  :  |  | | : :|  : : :   |||:|:  :: | ::||
Db        243 DVHWEFEQIQDNVRRQSSRVSWAEARAPHVCRPITVALLMRLLQQLTGITPILVYLQSIF 302

Qy        284 EEAKF----KDSSLASVVVGVIQVLFTAVAALIMDRAGRRLLLVLSGVVMVFSTSAFGAY 339
              :        ||    : :|| :::|   :||| || |||::|| :|  :|   :    | |
Db        303 DSTAVLLPPKDD---AAIVGAVRLLSVLIAALTMDLAGRKVLLFVSAAIMFAANLTLGLY 359

Qy        340 FKLTQGGPGNSSHVAISAPVS------AQPVDASVG-LAWLAVGSMCLFIAGFAVGWGPI 392
                    ||   |  : :   |      |||: |  | |  : : :  ||| |:|||||||
Db        360 IHF---GPRPLSPNSTAGLESESWGDLAQPLAAPAGYLTLVPLLATMLFIMGYAVGWGPI 416

Qy        393 PWLLMSEIFPLHVKGVATGICVLTNWLMAFLVTKEFSSLMEVLRPYGAFWLASAFCIFSV 452
               ||||||: ||  :|||:|:||| :|| ||::|| |  ::        |:  :| |: |:
Db        417 TWLLMSEVLPLRARGVASGLCVLASWLTAFVLTKSFLPVVSTFGLQVPFFFFAAICLVSL 476

Qy        453 LFTLFCVPETKGKTLEQITAHFE-GR 477
              :||  |||||||::|||| : |  ||
Db        477 VFTGCCVPETKGRSLEQIESFFRTGR 502
```

```
RESULT 14
ADA84077
ID   ADA84077 standard; protein; 507 AA.
XX
AC   ADA84077;
XX
DT   20-NOV-2003  (first entry)
XX
DE   Human SLC2A6 protein.
XX
KW   human; marker; expressed sequence tag; EST; arabidopsis; tumour;
KW   stress-induced phenotype; hyperosmotic stress; colon cancer; immunogen;
KW   vaccine.
XX
OS   Homo sapiens.
XX
PN   WO2002103028-A2.
XX
PD   27-DEC-2002.
XX
PF   30-MAY-2002; 2002WO-IB004189.
XX
PR   30-MAY-2001; 2001US-0293999P.
PR   22-OCT-2001; 2001US-0330457P.
PR   19-FEB-2002; 2002US-0357144P.
XX
PA   (BIOM-) BIOMEDICAL CENT.
XX
PI   Baranova AV,  Yankovsky NK,  Kozlov AP,  Lobashev AV,  Krukovskaya LL;
XX
DR   WPI; 2003-175241/17.
DR   N-PSDB; ADA84076.
XX
PT   Determining if a nucleic acid is a marker for a phenotype/cell type of
PT   interest, by global comparison of expressed sequence tags known to be
PT   expressed in the phenotype/cell type with all ESTs expressed in normal
PT   tissue.
XX
PS   Claim 29; Page 458-460; 516pp; English.
XX
CC   The invention relates to a novel method for determining if a nucleic acid
CC   is a marker for a predetermined phenotype/cell type of interest from a
CC   biological species. The method comprises performing a global comparison
CC   of a group of expressed sequence tags (ESTs) known to be expressed in the
CC   phenotype/cell type of interest with all ESTs expressed in normal tissue
CC   in order to identify ESTs that are preferentially expressed in the
CC   phenotype/cell of interest. A method of the invention is useful for
CC   determining whether a nucleic acid is a marker for a predetermined
CC   phenotype or cell type of interest from a biological species, preferably
CC   Arabidopsis or human. The cell type of interest is an abnormal cell such
CC   as a tumour cell, and the predetermined phenotype is a stress-induced
CC   phenotype such as hyperosmotic stress or high salt conditions. A method
CC   of the invention is also useful for determining the progression of colon
CC   cancer in a human, for detecting a tumour cell, and for regulating or
CC   preventing the growth of a tumour cell. An antibody of the invention is
CC   useful for detecting the absence or presence of peptides encoded by
CC   tumour-associated markers. A polypeptide of the invention is useful as an
CC   immunogen for vaccinating an animal. The present sequence represents a
CC   tumour-associated antigen of the invention.
XX
SQ   Sequence 507 AA;

  Query Match             38.6%;  Score 948;  DB 6;  Length 507;
  Best Local Similarity   42.7%;  Pred. No. 4.9e-82;
  Matches  216;  Conservative   80;  Mismatches  166;  Indels   44;  Gaps   11;

Qy         10 QPLLG------------PPGGSAPRGR-------RVFLAAFAAALGPLSFGFALGYSSPA 50
              :||||            ||     | |       ||||| ||| ||  |||:|| |:||
Db          3 EPLLGAEGPDYDTFPEKPPPSPGDRARVGTLQNKRVFLATFAAVLGNFSFGYALVYTSPV 62

Qy         51 IPSLQRAAPPAPRLDDAAASWFGAVVTLGAAAGGVLGGWLVDRAGRKLSLLLCSVPFVAG 110
              ||:|:|:  |   |  : |||||:| ||||||||:    | |  |||||::  :||  ||
Db         63 IPALERSLDPDLHLTKSQASWFGSVFTLGAAAGGLSAMILNDLLGRKLSIMFSAVPSAAG 122

Qy        111 FAVITAAQDVWMLLGGRLLTGLACGVASLVAPVYISEIAYPAVRGLLGSCVQLMVVVGIL 170
              :|::  |  :|||| || ||| | |: :   |||:|||| | ||| ||:  ||| | | |
Db        123 YALMAGAHGLWMLLLGRTLTGFAGGLTAACIPVYVSEIAPPGVRGALGATPQLMAVFGSL 182

Qy        171 LAYLAGWVLEWRWLAVLGCVPPSLMLLLMCFMPETPRFLLTQHRRQEAMAALRFLWGSEQ 230
                |  | :| |||||| |  |  :|:||: ||| :|||||:: | :||: || :| |::
Db        183 SLYALGLLLPWRWLAVAGEAPVLIMILLLSFMPNSPRFLLSRGRDEEALRALAWLRGTDV 242

Qy        231 G--WEDPPIG---AEQSFHL--ALLRQPGIYKPFIIGVSLMAFQQLSGVNAVMFYAETIF 283
                 ||   |      ||  :  | | | : :|  : : :   |||:|:  :: | ::||
Db        243 DVHWEFEQIQDNVRRQSSRVSWAEARAPHVCRPITVALLMRLLQQLTGITPILVYLQSIF 302

Qy        284 EEAKF----KDSSLASVVVGVIQVLFTAVAALIMDRAGRRLLLVLSGVVMVFSTSAFGAY 339
              :        ||    : :|| :::|   :||| || |||::|| :|  :|  :    | |
Db        303 DSTAVLLPPKDD---AAIVGAVRLLSVLIAALTMDLAGRKVLLFVSAAIMFAANLTLGLY 359

Qy        340 FKLTQGGPGNSSHVAISAPVS------AQPVDASVG-LAWLAVGSMCLFIAGFAVGWGPI 392
                    ||   |  : :   |      |||: |  | |  : : :  ||| |:|||||||
Db        360 IHF---GPRPLSPNSTAGLESESWGDLAQPLAAPAGYLTLVPLLATMLFIMGYAVGWGPI 416

Qy        393 PWLLMSEIFPLHVKGVATGICVLTNWLMAFLVTKEFSSLMEVLRPYGAFWLASAFCIFSV 452
               ||||||: ||  :|||:|:||| :|| ||::|| |  ::        |:  :| |: |:
Db        417 TWLLMSEVLPLRARGVASGLCVLASWLTAFVLTKSFLPVVSTFGLQVPFFFFAAICLVSL 476

Qy        453 LFTLFCVPETKGKTLEQITAHFE-GR 477
              :||  |||||||::|||| : |  ||
Db        477 VFTGCCVPETKGRSLEQIESFFRTGR 502


RESULT 15
ABP58365
ID   ABP58365 standard; protein; 507 AA.
XX
AC   ABP58365;
XX
DT   07-APR-2003  (first entry)
XX
DE   Human solute carrier type 2A polypeptide 8923733.
XX
KW   Human; solute carrier type 2A; SLC2A; glucose transporter; p53; cancer;
KW   cytostatic; gene therapy.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Domain          39. .61
FT                   /note= "transmembrane domain"
FT   Domain          40. .498
FT                   /note= "sugar transporter domain"
FT   Domain          81. .103
FT                   /note= "transmembrane domain"
FT   Domain          110. .132
FT                   /note= "transmembrane domain"
FT   Domain          136. .158
FT                   /note= "transmembrane domain"
FT   Domain          165. .187
FT                   /note= "transmembrane domain"
FT   Domain          191. .213
FT                   /note= "transmembrane domain"
FT   Domain          279. .301
FT                   /note= "transmembrane domain"
FT   Domain          316. .333
FT                   /note= "transmembrane domain"
FT   Domain          340. .362
FT                   /note= "transmembrane domain"
FT   Domain          391. .413
FT                   /note= "transmembrane domain"
FT   Domain          426. .448
FT                   /note= "transmembrane domain"
FT   Domain          458. .480
```

```
FT                   /note= "transmembrane domain"
XX
PN   WO200298467-A1.
XX
PD   12-DEC-2002.
XX
PF   03-JUN-2002; 2002WO-US017419.
XX
PR   05-JUN-2001; 2001US-0296076P.
PR   10-OCT-2001; 2001US-0328605P.
PR   15-FEB-2002; 2002US-0357253P.
XX
PA   (EXEL-) EXELIXIS INC.
XX
PI   Friedman L,  Plowman GD,  Belvin M,  Francis-Lang H,  Li D,  Funke RP;
XX
DR   WPI; 2003-201283/19.
XX
PT   Identifying candidate p53 pathway modulating agent as therapeutic target
PT   for disorders of defective p53 function e.g. cancer, by assaying purified
PT   solute carrier family 2 (SLC2A) polypeptide or nucleic acid with a test
PT   agent.
XX
PS   Claim 13; Page 48-50; 58pp; English.
XX
CC   The present sequence is that of human solute carrier type 2A (SLC2A)
CC   polypeptide 89237334. In the present invention, genetic screens were
CC   designed to identify modifiers of the p53 pathway in Drosophila in which
CC   p53 was overexpressed. Human orthologues (polynucleotides and
CC   polypeptides) of one such modifier were then identified, including the
CC   present polypeptide. SLC2As are glucose transporter proteins with sugar
CC   transporter domains. They are attractive drug targets for the treatment
CC   of pathologies associated with a defective p53 signalling pathway, such
CC   as cancer. The invention provides in vitro and in vivo methods of
CC   assessing SLC2A function. Modulation of an SLC2A or its binding partners
CC   is useful for understanding the association of the p53 pathway and its
CC   members in normal and disease conditions and for developing diagnostic
CC   and therapeutic modalities of p53-related pathologies. SLC2A-modulating
CC   agents that act by inhibiting or enhancing SLC2A expression, directly or
CC   indirectly, e.g. by affecting an SLC2A function such as transport or
CC   binding activity, can be identified using methods provided. Modulators
CC   include small molecules, nucleic acids, antibodies, antisense
CC   oligonucleotides and phosphothioate morpholino oligomers (claimed)
XX
SQ   Sequence 507 AA;

  Query Match             38.6%;  Score 948;  DB 6;  Length 507;
  Best Local Similarity   42.7%;  Pred. No. 4.9e-82;
  Matches  216;  Conservative   80;  Mismatches  166;  Indels   44;  Gaps   11;

Qy         10 QPLLG------------PPGGSAPRGR-------RVFLAAFAAALGPLSFGFALGYSSPA 50
              :||||            ||     | |       ||||| ||| ||  |||:|| |:||
Db          3 EPLLGAEGPDYDTFPEKPPPSPGDRARVGTLQNKRVFLATFAAVLGNFSFGYALVYTSPV 62

Qy         51 IPSLQRAAPPAPRLDDAAASWFGAVVTLGAAAGGVLGGWLVDRAGRKLSLLLCSVPFVAG 110
              ||:|:|:  |   |  : |||||:| ||||||||:    | |  |||||::  :||  ||
Db         63 IPALERSLDPDLHLTKSQASWFGSVFTLGAAAGGLSAMILNDLLGRKLSIMFSAVPSAAG 122

Qy        111 FAVITAAQDVWMLLGGRLLTGLACGVASLVAPVYISEIAYPAVRGLLGSCVQLMVVVGIL 170
              :|::  |  :|||| || ||| | |: :   |||:|||| | ||| ||:  ||| | | |
Db        123 YALMAGAHGLWMLLLGRTLTGFAGGLTAACIPVYVSEIAPPGVRGALGATPQLMAVFGSL 182

Qy        171 LAYLAGWVLEWRWLAVLGCVPPSLMLLLMCFMPETPRFLLTQHRRQEAMAALRFLWGSEQ 230
                |  | :| |||||| |  |  :|:||: ||| :|||||:: | :||: || :| |::
Db        183 SLYALGLLLPWRWLAVAGEAPVLIMILLLSFMPNSPRFLLSRGRDEEALRALAWLRGTDV 242

Qy        231 G--WEDPPIG---AEQSFHL--ALLRQPGIYKPFIIGVSLMAFQQLSGVNAVMFYAETIF 283
                 ||   |      ||  :  |  | | : :|  : : :   |||:|:  :: | ::||
Db        243 DVHWEFEQIQDNVRRQSSRVSWAEARAPHVCRPITVALLMRLLQQLTGITPILVYLQSIF 302

Qy        284 EEAKF----KDSSLASVVVGVIQVLFTAVAALIMDRAGRRLLLVLSGVVMVFSTSAFGAY 339
              :        ||    : :|| :::|   :||| || |||::|| :|  :|  :    | |
Db        303 DSTAVLLPPKDD---AAIVGAVRLLSVLIAALTMDLAGRKVLLFVSAAIMFAANLTLGLY 359

Qy        340 FKLTQGGPGNSSHVAISAPVS------AQPVDASVG-LAWLAVGSMCLFIAGFAVGWGPI 392
                    ||   |  : :   |      |||: |  | |  : : :  ||| |:|||||||
Db        360 IHF---GPRPLSPNSTAGLESESWGDLAQPLAAPAGYLTLVPLLATMLFIMGYAVGWGPI 416

Qy        393 PWLLMSEIFPLHVKGVATGICVLTNWLMAFLVTKEFSSLMEVLRPYGAFWLASAFCIFSV 452
               ||||||: ||  :|||:|:||| :|| ||::|| |  ::        |:  :| |: |:
Db        417 TWLLMSEVLPLRARGVASGLCVLASWLTAFVLTKSFLPVVSTFGLQVPFFFFAAICLVSL 476

Qy        453 LFTLFCVPETKGKTLEQITAHFE-GR 477
              :||  |||||||::|||| : |  ||
Db        477 VFTGCCVPETKGRSLEQIESFFRMGR 502

Search completed: September 27, 2004, 19:26:41
Job time : 130 secs
```

```
                       GenCore version 5.1.6
              Copyright (c) 1993 - 2004  Compugen Ltd.

OM protein - protein search, using sw model

Run on:         September 27, 2004, 19:22:56 ; Search time 45 Seconds
                                           (without alignments)
                                           1019.630 Million cell updates/sec

Title:          US-09-886-954A-1
Perfect score:  2457
Sequence:       1 MTPEDPEETQPLLGPPGGSA..........CVPETKGKTLEQITAHFEGR 477

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters:      283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       PIR_78:*
                1:  pir1:*
                2:  pir2:*
                3:  pir3:*
                4:  pir4:*

     Pred. No. is the number of results predicted by chance to have a
     score greater than or equal to the score of the result being printed,
     and is derived by analysis of the total score distribution.

                                 SUMMARIES
                  %
Result           Query
  No.    Score   Match Length DB  ID                   Description
-----------------------------------------------------------------------------
     1     734    29.9    487  2  E96782               hypothetical prote
     2     722    29.4    490  2  T14545               probable sugar tra
     3     607    24.7    461  2  D70073               metabolite transpo
     4     607    24.7    471  2  AB0868               L-arabinose isomer
     5   606.5    24.7    472  2  S47089               arabinose-proton s
     6     603    24.5    472  2  B26430               L-arabinose isomer
     7     603    24.5    472  2  B91091               L-arabinose isomer
     8     603    24.5    472  2  E85936               L-arabinose isomer
     9   591.5    24.1    464  2  F65079               galactose-proton s
    10   587.5    23.9    464  2  C91106               galactose-proton s
    11   587.5    23.9    464  2  F85951               galactose-proton s
    12   584.5    23.8    496  2  T52132               probable sugar tra
    13   584.5    23.8    575  2  T43400               myo-inositol trans
    14   583.5    23.7    464  2  AC0877               galactose-proton s
    15     582    23.7    521  2  G84864               probable membrane
    16   574.5    23.4    493  2  S38981               glucose transport
    17   570.5    23.2    496  2  A31986               glucose transporte
    18   569.5    23.2    493  2  A41751               glucose transport
    19     567    23.1    580  2  D86426               hypothetical prote
    20   560.5    22.8    457  2  E70070               metabolite transpo
    21   556.5    22.6    560  2  T51485               sugar transporter-
    22   554.5    22.6    464  2  F69587               L-arabinose transp
    23     551    22.4    461  2  G85059               probable sugar tra
    24   548.5    22.3    523  2  S25015               monosaccharide tra
    25   546.5    22.2    526  2  T01853               probable hexose tr
    26     546    22.2    491  2  A26430               xylose transport p
    27     546    22.2    491  2  F91255               xylose-proton symp
    28     546    22.2    491  2  B86096               xylose-proton symp
    29   542.5    22.1    522  2  A31556               glucose transport
    30     542    22.1    534  2  S38435               hexose transport p
    31     538    21.9    523  2  S06920               glucose transport
    32   536.5    21.8    516  2  T12199               monosaccharid tran
    33     536    21.8    509  2  A33801               muscle-fat glucose
    34   535.5    21.8    469  2  H97064               probable sugar-pro
    35   535.5    21.8    522  2  E86246               glucose transporte
    36     535    21.8    457  2  H85059               probable sugar tra
    37     535    21.8    514  2  S25009               monosaccharide tra
    38     533    21.7    547  2  C84593               probable sugar tra
    39   532.5    21.7    522  2  S12042               glucose transport
    40   531.5    21.6    496  2  A41264               glucose transport
    41   531.5    21.6    507  2  T01844               probable sugar tra
    42     531    21.6    523  2  T10122               hexose transport p
    43   530.5    21.6    468  2  S10014               glucose transport
    44     530    21.6    509  2  A32101               glucose transport
    45   528.5    21.5    534  2  S14144               hexose transport p


                                 ALIGNMENTS


RESULT 1
E96782
hypothetical protein F22H5.6 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001
C;Accession: E96782
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: E96782
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-487 <STO>
A;Cross-references: GB:AE005173; NID:g10092276; PIDN:AAG12689.1; GSPDB:GN00141
C;Genetics:
A;Gene: F22H5.6
A;Map position: 1
C;Superfamily: glucose transport protein

  Query Match             29.9%;  Score 734;  DB 2;  Length 487;
  Best Local Similarity   37.5%;  Pred. No. 8.9e-47;
  Matches  173;  Conservative   77;  Mismatches  177;  Indels   34;  Gaps    6;

Qy         26 VFLAAFAAALGPLSFGFALGYSSPAIPSLQRAAPPAPRLDDAAASWFGAVVTLGAAAGGV 85
              |       ||||: |||  |||||     | |      |  :  | ||::  :||  | :
Db         48 VLACVLIVALGPIQFGFTCGYSSPT----QAAITKDLGLTVSEYSVFGSLSNVGAMVGAI 103

Qy         86 LGGWLVDRAGRKLSLLLCSVPFVAGFAVITAAQDVWMLLGGRLLTGLACGVASLVAPVYI 145
                | : :  ||| ||:: ::| : |:  |: |:|   |  |||| |   |: |   ||||
Db        104 ASGQIAEYIGRKGSLMIAAIPNIIGWLCISFAKDTSFLYMGRLLEGFGVGIISYTVPVYI 163

Qy        146 SEIAYPAVRGLLGSCVQLMVVVGILLAYLAGWVLEWRWLAVLGCVPPSLMLLLMCFMPET 205
              :|||   :|| |||  || | :||:|||| |  : || ||||| :| :|::  : |:||:
Db        164 AEIAPQNMRGGLGSVNQLSVTIGIMLAYLLGLFVPWRILAVLGILPCTLLIPGLFFIPES 223

Qy        206 PRFLLTQHRRQEAMAALRFLWGSEQGWEDPPIGAEQ------------SFHLALLRQPGI 253
              ||:|       |   :|: | | |    |  :   :            :     |::
Db        224 PRWLAKMGMTDEFETSLQVLRGFE---TDITVEVNEIKRSVASSTKRNTVRFVDLKRRRY 280

Qy        254 YKPFIIGVSLMAFQQLSGVNAVMFYAETIFEEAKFKDSSLASVVVGVIQVLFTAVAALIM 313
              | | ::|: |:  ||| |:| |:||: |||| |    |: |:  || |||: ||::  ::
Db        281 YFPLMVGIGLLVLQQLGGINGVLFYSSTIFESAGVTSSNAATFGVGAIQVVATAISTWLV 340

Qy        314 DRAGRRLLLVLSGVVMVFSTSAFGAYFKLTQGGPGNSSHVAISAPVSAQPVDASVGLAWL 373
```

```
               |:||||||| :| | |  |     | | | :            :| |    |    |: |
Db         341 DKAGRRLLLTISSVGMTISLVIVAAAFYLKE----------FVSPDS----DMYSWLSIL 386

Qy         374 AVGSMCLFIAGFAVGWGPIPWLLMSEIFPLHVKGVATGICVLTNWLMAFLVTKEFSSLME 433
               :|  :   :  |::| ||||||:|||| |:::||:|  |  | ||  ::|:|   ::|:
Db         387 SVVGVVAMVVFFSLGMGPIPWLIMSEILPVNIKGLAGSIATLANWFFSWLITMT-ANLLL 445

Qy         434 VLRPYGAFWLASAFCIFSVLFTLFCVPETKGKTLEQITAHF 474
                    | | |    | |:|:|    ||||||||||:: : |
Db         446 AWSSGGTFTLYGLVCAFTVVFVTLWVPETKGKTLEELQSLF 486

RESULT 2
T14545
probable sugar transporter protein - beet
C;Species: Beta vulgaris (beet)
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 21-Jul-2000
C;Accession: T14545
R;Chiou, T.J.; Bush, D.R.
Plant Physiol. 110, 511-520, 1996
A;Title: Molecular cloning, immunochemical localization to the vacuole, and expression i
A;Reference number: Z18131; MUID:96351183; PMID:8742332
A;Accession: T14545
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-490 <CHI>
A;Cross-references: EMBL:U43629; NID:g1209755; PIDN:AAB53155.1; PID:g1209756
A;Experimental source: tonoplast
C;Superfamily: glucose transport protein
C;Keywords: transmembrane protein

  Query Match             29.4%;  Score 722;  DB 2;  Length 490;
  Best Local Similarity   36.1%;  Pred. No. 6.9e-46;
  Matches  167;  Conservative   81;  Mismatches  177;  Indels   38;  Gaps    5;

Qy          26 VFLAAFAAALGPLSFGFALGYSSPAIPSLQRAAPPAPRLDDAAASWFGAVVTLGAAAGGV 85
               |       ||||: |||  |||||     | |      |  |  ||||::  :||  | :
Db          50 VLACVLIVALGPIQFGFTAGYSSPT----QSAITNELGLSVAEYSWFGSLSNVGAMVGAI 105

Qy          86 LGGWLVDRAGRKLSLLLCSVPFVAGFAVITAAQDVWMLLGGRLLTGLACGVASLVAPVYI 145
                 | : :  ||| ||:: ::| : |:  |: |:|   |  ||:| |    : |   ||||
Db         106 ASGQISEYIGRKGSLMIAAIPNIIGWLAISFAKDSSFLYMGRMLEGFGVGIISYTVPVYI 165

Qy         146 SEIAYPAVRGLLGSCVQLMVVVGILLAYLAGWVLEWRWLAVLGCVPPSLMLLLMCFMPET 205
               ||||   :|| |||  || | :||:|:|: |  : || ||||| :| ::::  : |:||:
Db         166 SEIAPQNLRGALGSVNQLSVTIGIMLSYMLGLFVPWRILAVLGILPCTILIPGLFFIPES 225

Qy         206 PRFLLTQHRRQEAMAALRFLWGSEQGWEDPPIGAE--------------QSFHLALLRQP 251
               ||:|      :|   :|: | |      |  |  |               :   | |||
Db         226 PRWLAKMGMMEEFEVSLQVLRGF-----DTDISLEVNEIKRSVASSSKRTTIRFAELRQR 280

Qy         252 GIYKPFIIGVSLMAFQQLSGVNAVMFYAETIFEEAKFKDSSLASVVVGVIQVLFTAVAAL 311
                 : | :||  |:  |||||:| |:||: |||:||    |: |:  :| :||: | |
Db         281 RYWLPLMIGNGLLILQQLSGINGVLFYSSTIFKEAGVTSSNAATFGLGAVQVIATVVTTW 340

Qy         312 IMDRAGRRLLLVLSGVVMVFSTSAFGAYFKLTQGGPGNSSHVAISAPVSAQPVDASVGLA 371
               ::|::||||||::|   |  |       | | :     |:  ::               :
Db         341 LVDKSGRRLLLIVSSSGMTLSLLVVAMSFFLKEMVSDESTWYSV--------------FS 386

Qy         372 WLAVGSMCLFIAGFAVGWGPIPWLLMSEIFPLHVKGVATGICVLTNWLMAFLVTKEFSSL 431
                |:|  :   :  |::| | |||::|||| |:::||:|  |  | || :|::||   :::
Db         387 ILSVVGVVAMVVTFSLGIGAIPWIIMSEILPINIKGLAGSIATLANWFVAWIVTMT-ANI 445

Qy         432 MEVLRPYGAFWLASAFCIFSVLFTLFCVPETKGKTLEQITAHF 474
               |      | | :    | |:| | :  ||||||:|||:|   |
Db         446 MLSWNSGGTFSIYMVVCAFTVAFVVIWVPETKGRTLEEIQWSF 488

RESULT 3
D70073
metabolite transport protein homolog yxcC - Bacillus subtilis
C;Species: Bacillus subtilis
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Jun-2000
```

```
C;Accession: D70073
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berte
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galle
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mauee
y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetell
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Sero
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida,
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A;Reference number: A69580; MUID:98044033; PMID:9384377
A;Accession: D70073
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-461 <KUN>
A;Cross-references: GB:Z99124; GB:AL009126; NID:g2636442; PIDN:CAB16017.1; PID:g2636527
A;Experimental source: strain 168
C;Genetics:
A;Gene: yxcC
C;Superfamily: glucose transport protein

  Query Match             24.7%;  Score 607;  DB 2;  Length 461;
  Best Local Similarity   31.1%;  Pred. No. 2e-37;
  Matches  146;  Conservative   85;  Mismatches  192;  Indels   46;  Gaps    7;

Qy          25 RVFLAAFAAALGPLSFGFALGYSSPAIPSLQRAAPPAPRLDDAAASWFGAVVTLGAAAGG 84
               | ::  |  ||| | :|:  |  | |:  :    |     :    |    :: |||  |
Db           6 RKYMIYFFGALGGLLYGYDTGVISGALLFINNDIPLTTLTEGLVVS----MLLLGAIFGS 61

Qy          85 VLGGWLVDRAGRKLSLLLCSVPFVAGFAVITAAQDVWMLLGGRLLTGLACGVASLVAPVY 144
                | |   || ||:  : : |: |: |      :| : ||:  |:: ||| | :: : |||
Db          62 ALSGTCSDRWGRRKVVFVLSIIFIIGALACAFSQTIGMLIASRVILGLAVGGSTALVPVY 121

Qy         145 ISEIAYPAVRGLLGSCVQLMVVVGILLAYLAGWVL----EWRWLAVLGCVPPSLMLLLMC 200
               :||:|   :|| ||:   ||:| ||||||:  ::      |||:  |  ||  |:|: :
Db         122 LSEMAPTKIRGTLGTMNNLMIVTGILLAYIVNYLFTPFEAWRWMVGLAAVPAVLLLIGIA 181

Qy         201 FMPETPRFLLTQHRRQEAMAALRF----------LWGSEQGWEDPPIGAEQSFHLALLRQ 250
               ||||:||:|: :   :||   :            |   :||  :     ::   | :|:
Db         182 FMPESPRWLVKRGSEEEARRIMNITHDPKDIEMELAEMKQGEAE-----KKETTLGVLKA 236

Qy         251 PGIYKPFIIGVSLMAFQQLSGVNAVMFYAETIFEEAKF--KDSSLASVVVGVIQVLFTAV 308
                 |    :||| |  |||  |:| |::|| ||| :|      |:| :: :|:: |:
Db         237 KWIRPMLLIGVGLAIFQQAVGINTVIYYAPTIFTKAGLGTSASALGTMGIGILNVIMCIT 296

Qy         309 AALIMDRAGRRLLLVLSGVVMVFSTSAFGAYFKLTQGGPGNSSHVAISAPVSAQPVDASV 368
               | :::|| ||: ||:   | :  | :|      || |                     |
Db         297 AMILIDRVGRKKLLIWGSVGITLSLAALSGVL-LTLG--------------------LSA 335

Qy         369 GLAWLAVGSMCLFIAGFAVGWGPIPWLLMSEIFPLHVKGVATGICVLTNWLMAFLVTKEF 428
                 ||: |  : ::|  :   |||: |:|| |:||   :| |||   |       :|:  |
Db         336 STAWMTVVFLGVYIVFYQATWGPVVWVLMPELFPSKARGAATGFTTLVLSAANLIVSLVF 395

Qy         429 SSLMEVLRPYGAFWLASAFCIFSVLFTLFCVPETKGKTLEQITAHFEGR 477
                 ::  :     | : |  |: |  |  : |||||||:||:| | :  |
Db         396 PLMLSAMGIAWVFMVFSVICLLSFFFAFYMVPETKGKSLEEIEASLKKR 444

RESULT 4
AB0868
L-arabinose isomerase [imported] - Salmonella enterica subsp. enterica serovar Typhi (st
C;Species: Salmonella enterica subsp. enterica serovar Typhi
A;Note: this species has also been called Salmonella typhi
C;Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C;Accession: AB0868
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
```

```
, S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A;Reference number: AB0502; MUID:21534947; PMID:11677608
A;Accession: AB0868
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-471 <PAR>
A;Cross-references: GB:AL513382; PIDN:CAD02842.1; PID:g16504096; GSPDB:GN00176
C;Genetics:
A;Gene: STY3160
C;Superfamily: glucose transport protein

  Query Match             24.7%;  Score 607;  DB 2;  Length 471;
  Best Local Similarity   32.7%;  Pred. No. 2.1e-37;
  Matches  158;  Conservative   85;  Mismatches  192;  Indels   48;  Gaps   10;

Qy          1 MTPEDPEETQPLLGPPGGSAPRGRRVFLAAFAAALGPLSFGFALGYSSPAIPSLQRAAPP 60
              :||    :|            |   :|::  ||  | | ||  :|  : |:| :
Db          9 LTPRSLRDT------------RRMNMFVSVSAAVAG-LLFGLDIGVIAGALPFITDHFVL 55

Qy         61 APRLDDAAASWFGAVVTLGAAAGGVLGGWLVDRAGRKLSLLLCSVPFVAGFAVITAAQDV 120
                || :    |  : : ||||| | :  |||  | ||| ||:  :: || |      |  |
Db         56 TSRLQE----WVVSSMMLGAAIGALFNGWLSFRLGRKYSLMAGAILFVLGSLGSAFASSV 111

Qy        121 WMLLGGRLLTGLACGVASLVAPVYISEIAYPAVRGLLGSCVQLMVVVGILLAYLAGWVLE 180
               :|:| |:: |:| |:||  ||:|:||:|   ||| : |  |||| :||:||:|:
Db        112 EVLIGARVILGVAVGIASYTAPLYLSEMASENVRGKMISMYQLMVTLGIVLAFLSDTAFS 171

Qy        181 W--RWLAVLG--CVPPSLMLLLMCFMPETPRFLLTQHRRQEAMAALRFLWGSEQGWEDPP 236
              :   | |:||   :|  |:::|: |:| :||:|  : |  ||   || |  : :   |
Db        172 YSGNWRAMLGVLALPAVLLIILVVFLPNSPRWLAQKGRHIEAEEVLRMLRDTSEKARDEL 231

Qy        237 IGAEQSFHL-----ALLR-QPGIYKPFIIGVSLMAFQQLSGVNAVMFYAETIFEEAKF-- 288
                  :|  |     || :    : :   :|: | | || :|:| :|:||  ||: | |
Db        232 NEIRESLKLKQGGWALFKANRNVRRAVFLGMLLQAMQQFTGMNIIMYYAPRIFKMAGFTT 291

Qy        289 -KDSSLASVVVGVIQVLFTAVAALIMDRAGRRLLLVLSGVVMVFSTSAFGAYFKLTQGGP 347
               :   :|::|||:  :  | :|   :|:|||:  | :   ||   |   |  : | |
Db        292 TEQQMIATLVVGLTFMFATFIAVFTVDKAGRKPALKIGFSVMALGTLVLG--YCLMQFDN 349

Qy        348 GNSSHVAISAPVSAQPVDASVGLAWLAVGSMCLFIAGFAVGWGPIPWLLMSEIFPLHVKG 407
              |                 || ||:||:||   : |||:|:   |: |:| ||| ||  :
Db        350 GT----------------ASSGLSWLSVGMTMMCIAGYAMSAAPVVWILCSEIQPLKCRD 393

Qy        408 VATGICVLTNWLMAFLVTKEFSSLMEVLRPYGAFWLASAFCIFSVLFTLFCVPETKGKTL 467
                      |||:   ::   | :|:: :   | ||| :|  |  :  | : :||||  ||
Db        394 FGITCSTTTNWVSNMIIGATFLTLLDSIGAAGTFWLYTALNIAFIGITFWLIPETKNVTL 453

Qy        468 EQI 470
              | |
Db        454 EHI 456

RESULT 5
S47089
arabinose-proton symporter - Klebsiella oxytoca
C;Species: Klebsiella oxytoca
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 24-Sep-1999
C;Accession: S47089
R;Shatwell, K.P.; Charalambous, B.M.; McDonald, T.P.; Henderson, P.J.F.
submitted to the EMBL Data Library, June 1994
A;Description: The nucleotide sequence of the gene araE for arabinose-proton synport in
A;Reference number: S47089
A;Accession: S47089
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-472 <SHA>
A;Cross-references: EMBL:X79598; NID:g498919; PIDN:CAA56110.1; PID:g498920
C;Superfamily: glucose transport protein
C;Keywords: arabinose transport
```

```
  Query Match             24.7%;  Score 606.5;  DB 2;  Length 472;
  Best Local Similarity   33.7%;  Pred. No. 2.3e-37;
  Matches  153;  Conservative   84;  Mismatches  182;  Indels   35;  Gaps    8;

Qy         30 AFAAALGPLSFGFALGYSSPAIPSLQRAAPPAPRLDDAAASWFGAVVTLGAAAGGVLGGW 89
              : |||:  | ||  :|  : |:| :      : || :    |  : : |||| | :  ||
Db         26 SIAAAVAGLLFGLDIGVIAGALPFITDHFVLSSRLQE----WVVSSMMLGAAIGALFNGW 81

Qy         90 LVDRAGRKLSLLLCSVPFVAGFAVITAAQDVWMLLGGRLLTGLACGVASLVAPVYISEIA 149
              |  | ||| ||:: :| |||||      |  | |||  |:: |:| |:||  ||:|:||:|
Db         82 LSFRLGRKYSLMVGAVLFVAGSVGSAFATSVEMLLVARIVLGVAVGIASYTAPLYLSEMA 141

Qy        150 YPAVRGLLGSCVQLMVVVGILLAYLAGWVLEW--RWLAVLG--CVPPSLMLLLMCFMPET 205
                 ||| : |  ||||:|||::|:|      :   | |:||   :|  ::::|: |:| :
Db        142 SENVRGKMISMYQLMVTLGIVMAFLSDTAFSYSGNWRAMLGVLALPAVVLIIILVIFLPNS 201

Qy        206 PRFLLTQHRRQEAMAALRFLWGSEQGWEDPPIGAEQSFHL-----ALLR-QPGIYKPFII 259
              ||:|  : |  ||   || |  : :   |      :|  |     || :    : :   :
Db        202 PRWLAEKGRHVEAEEVLRMLRDTSEKARDELNEIRESLKLKQGGWALFKVNRNVRRAVFL 261

Qy        260 GVSLMAFQQLSGVNAVMFYAETIFEEAKF---KDSSLASVVVGVIQVLFTAVAALIMDRA 316
              |: | | || :|:| :|:||  ||: | |   :   :|::|||:  :  | :|   :|:|
Db        262 GMLLQAMQQFTGMNIIMYYAPRIFKMAGFTTTEQQMVATLVVGLTFMFATFIAVFTVDKA 321

Qy        317 GRRLLLVLSGVVMVFSTSAFGAYFKLTQGGPGNSSHVAISAPVSAQPVDASVGLAWLAVG 376
              ||:  | :   ||   |   |  : | |   |                 || ||:||:||
Db        322 GRKPALKIGFSVMAIGTLVLG--YCLMQFDNGT----------------ASSGLSWLSVG 363

Qy        377 SMCLFIAGFAVGWGPIPWLLMSEIFPLHVKGVATGICVLTNWLMAFLVTKEFSSLMEVLR 436
                 : |||:|:   |: |:| ||| ||  :         |||:   ::   | :|:: :
Db        364 MTMMCIAGYAMSAAPVVWILCSEIQPLKCRDFGITCSTTTNWVSNMIIGATFLTLLDAIG 423

Qy        437 PYGAFWLASAFCIFSVLFTLFCVPETKGKTLEQI 470
                | ||| :|  :  :  | : :||||  ||| |
Db        424 AAGTFWLYTALNVAFIGVTFWLIPETKNVTLEHI 457

RESULT 6
B26430
L-arabinose isomerase (EC 5.3.1.4) - Escherichia coli (strain K-12)
C;Species: Escherichia coli
C;Date: 05-Oct-1988 #sequence_revision 05-Oct-1988 #text_change 01-Mar-2002
C;Accession: B26430; A28075; I40996; B65067
R;Maiden, M.C.J.; Davis, E.O.; Baldwin, S.A.; Moore, D.C.M.; Henderson, P.J.F.
Nature 325, 641-643, 1987
A;Title: Mammalian and bacterial sugar transport proteins are homologous.
A;Reference number: A93389; MUID:87115869; PMID:3543693
A;Accession: B26430
A;Status: nucleic acid sequence not shown; not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-472 <MAI>
R;Maiden, M.C.J.; Jones-Mortimer, M.C.; Henderson, P.J.F.
J. Biol. Chem. 263, 8003-8010, 1988
A;Title: The cloning, DNA sequence, and overexpression of the gene araE coding for arabi
A;Reference number: A28075; MUID:88228015; PMID:2836407
A;Accession: A28075
A;Molecule type: DNA
A;Residues: 1-472 <MA2>
A;Cross-references: GB:J03732; NID:g145320; PIDN:AAA23469.1; PID:g145321
R;Stoner, C.; Schleif, R.
J. Mol. Biol. 171, 369-381, 1983
A;Title: The araE low affinity L-arabinose transport promoter. Cloning, sequence, transc
A;Reference number: I40996; MUID:84114868; PMID:6319708
A;Accession: I40996
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-25,'Y',349,'R' <RES>
A;Cross-references: EMBL:X00272; NID:g40940; PIDN:CAA25075.1; PID:g40941
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617; PMID:9278503
```

```
A;Accession: B65067
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-472 <BLAT>
A;Cross-references: GB:AE000368; GB:U00096; NID:g2367165; PIDN:AAC75880.1; PID:g1789207;
A;Experimental source: strain K-12, substrain MG1655
C;Genetics:
A;Gene: araE
A;Map position: 61 min
C;Superfamily: glucose transport protein
C;Keywords: arabinose transport; intramolecular oxidoreductase; isomerase; membrane prot

  Query Match             24.5%;  Score 603;  DB 2;  Length 472;
  Best Local Similarity   33.3%;  Pred. No. 4.1e-37;
  Matches  158;  Conservative   84;  Mismatches  197;  Indels   36;  Gaps    9;

Qy          9 TQPLLGPPGGSAPRGRRVFLAAFAAALGPLSFGFALGYSSPAIPSLQRAAPPAPRLDDAA 68
              |:  | |      |   :|::  |||:  | ||  :|  : |:| :        || :
Db          6 TESALTPRSLRDTRRMNMFVSV-AAAVAGLLFGLDIGVIAGALPFITDHFVLTSRLQE-- 62

Qy         69 ASWFGAVVTLGAAAGGVLGGWLVDRAGRKLSLLLCSVPFVAGFAVITAAQDVWMLLGGRL 128
                |  : : |||| | :  |||  | ||| ||:  :: || |      |  | ||:  |:
Db         63 --WVVSSMMLGAAIGALFNGWLSFRLGRKYSLMAGAILFVLGSIGSAFATSVEMLIAARV 120

Qy        129 LTGLACGVASLVAPVYISEIAYPAVRGLLGSCVQLMVVVGILLAYLAGWVLEW--RWLAV 186
              : |:| |:||  ||:|:||:|   ||| : |  |||| :||:||:|:     :   | |:
Db        121 VLGIAVGIASYTAPLYLSEMASENVRGKMISMYQLMVTLGIVLAFLSDTAFSYSGNWRAM 180

Qy        187 LG--CVPPSLMLLLMCFMPETPRFLLTQHRRQEAMAALRFLWGSEQGWEDPPIGAEQSFH 244
              ||   :|  |:::|: |:| :||:|  : |  ||   || |  : :   :      :|
Db        181 LGVLALPAVLLIILVVFLPNSPRWLAEKGRHIEAEEVLRMLRDTSEKAREELNEIRESLK 240

Qy        245 L-----ALLR-QPGIYKPFIIGVSLMAFQQLSGVNAVMFYAETIFEEAKF---KDSSLAS 295
              |     || :    : :   :|: | | || :|:| :|:||  ||: | |   :   :|:
Db        241 LKQGGWALFKINRNVRRAVFLGMLLQAMQQFTGMNIIMYYAPRIFKMAGFTTTEQQMIAT 300

Qy        296 VVVGVIQVLFTAVAALIMDRAGRRLLLVLSGVVMVFSTSAFGAYFKLTQGGPGNSSHVAI 355
              :|||:  :  | :|   :|:|||:  | :   ||   |   |  : | |   |
Db        301 LVVGLTFMFATFIAVFTVDKAGRKPALKIGFSVMALGTLVLG--YCLMQFDNGT------ 352

Qy        356 SAPVSAQPVDASVGLAWLAVGSMCLFIAGFAVGWGPIPWLLMSEIFPLHVKGVATGICVL 415
                        || ||:||:||   : |||:|:   |: |:| ||| ||  :
Db        353 ----------ASSGLSWLSVGMTMMCIAGYAMSAAPVVWILCSEIQPLKCRDFGITCSTT 402

Qy        416 TNWLMAFLVTKEFSSLMEVLRPYGAFWLASAFCIFSVLFTLFCVPETKGKTLEQI 470
              |||:   ::   | :|:: :   | ||| :|  |  |  | : :||||  ||| |
Db        403 TNWVSNMIIGATFLTLLDSIGAAGTFWLYTALNIAFVGITFWLIPETKNVTLEHI 457

RESULT 7
B91091
L-arabinose isomerase (EC 5.3.1.4) - Escherichia coli (strain O157:H7, substrain RIMD 05
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 31-Dec-2001
C;Accession: B91091
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genc
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: B91091
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-472 <HAY>
A;Cross-references: GB:BA000007; PIDN:BAB37121.1; PID:g13363170; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs3698
C;Superfamily: glucose transport protein
C;Keywords: intramolecular oxidoreductase; isomerase

  Query Match             24.5%;  Score 603;  DB 2;  Length 472;
  Best Local Similarity   33.3%;  Pred. No. 4.1e-37;
  Matches  158;  Conservative   84;  Mismatches  197;  Indels   36;  Gaps    9;

Qy          9 TQPLLGPPGGSAPRGRRVFLAAFAAALGPLSFGFALGYSSPAIPSLQRAAPPAPRLDDAA 68
              |:  | |      |   :|::  |||:  | ||  :|  : |:| :        || :
Db          6 TESALTPRSLRDTRRMNMFVSV-AAAVAGLLFGLDIGVIAGALPFITDHFVLTSRLQE-- 62

Qy         69 ASWFGAVVTLGAAAGGVLGGWLVDRAGRKLSLLLCSVPFVAGFAVITAAQDVWMLLGGRL 128
                |  : : |||| | :  |||  | ||| ||:  :: || |      |  | ||:  |:
Db         63 --WVVSSMMLGAAIGALFNGWLSFRLGRKYSLMAGAILFVLGSIGSAFATSVEMLIAARV 120

Qy        129 LTGLACGVASLVAPVYISEIAYPAVRGLLGSCVQLMVVVGILLAYLAGWVLEW--RWLAV 186
              : |:| |:||  ||:|:||:|   ||| : |  |||| :||:||:|:     :   | |:
Db        121 VLGIAVGIASYTAPLYLSEMASENVRGKMISMYQLMVTLGIVLAFLSDTAFSYSGNWRAM 180

Qy        187 LG--CVPPSLMLLLMCFMPETPRFLLTQHRRQEAMAALRFLWGSEQGWEDPPIGAEQSFH 244
              ||   :|  |:::|: |:| :||:|  : |  ||   || |  : :   :      :|
Db        181 LGVLALPAVLLIILVVFLPNSPRWLAEKGRHIEAEEVLRMLRDTSEKAREELNEIRESLK 240

Qy        245 L-----ALLR-QPGIYKPFIIGVSLMAFQQLSGVNAVMFYAETIFEEAKF---KDSSLAS 295
              |     || :    : :   :|: | | || :|:| :|:||  ||: | |   :   :|:
Db        241 LKQGGWALFKINRNVRRAVFLGMLLQAMQQFTGMNIIMYYAPRIFKMAGFTTTEQQMIAT 300

Qy        296 VVVGVIQVLFTAVAALIMDRAGRRLLLVLSGVVMVFSTSAFGAYFKLTQGGPGNSSHVAI 355
              :|||:  :  | :|   :|:|||:  | :   ||   |   |  : | |   |
Db        301 LVVGLTFMFATFIAVFTVDKAGRKPALKIGFSVMALGTLVLG--YCLMQFDNGT------ 352

Qy        356 SAPVSAQPVDASVGLAWLAVGSMCLFIAGFAVGWGPIPWLLMSEIFPLHVKGVATGICVL 415
                        || ||:||:||   : |||:|:   |: |:| ||| ||  :
Db        353 ----------ASSGLSWLSVGMTMMCIAGYAMSAAPVVWILCSEIQPLKCRDFGITCSTT 402

Qy        416 TNWLMAFLVTKEFSSLMEVLRPYGAFWLASAFCIFSVLFTLFCVPETKGKTLEQI 470
              |||:   ::   | :|:: :   | ||| :|  |  |  | : :||||  ||| |
Db        403 TNWVSNMIIGATFLTLLDSIGAAGTFWLYTALNIAFVGITFWLIPETKNVTLEHI 457

RESULT 8
E85936
L-arabinose isomerase (EC 5.3.1.4) - Escherichia coli (strain O157:H7, substrain EDL933
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Dec-2001
C;Accession: E85936
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhev
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: E85936
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-472 <STO>
A;Cross-references: GB:AE005174; NID:g12517333; PIDN:AAG57953.1; GSPDB:GN00145; UWGP:Z41
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: araE
C;Superfamily: glucose transport protein
C;Keywords: intramolecular oxidoreductase; isomerase

  Query Match             24.5%;  Score 603;  DB 2;  Length 472;
  Best Local Similarity   33.3%;  Pred. No. 4.1e-37;
  Matches  158;  Conservative   84;  Mismatches  197;  Indels   36;  Gaps    9;

Qy          9 TQPLLGPPGGSAPRGRRVFLAAFAAALGPLSFGFALGYSSPAIPSLQRAAPPAPRLDDAA 68
              |:  | |      |   :|::  |||:  | ||  :|  : |:| :        || :
Db          6 TESALTPRSLRDTRRMNMFVSV-AAAVAGLLFGLDIGVIAGALPFITDHFVLTSRLQE-- 62

Qy         69 ASWFGAVVTLGAAAGGVLGGWLVDRAGRKLSLLLCSVPFVAGFAVITAAQDVWMLLGGRL 128
                |  : : |||| | :  |||  | ||| ||:  :: || |      |  | ||:  |:
Db         63 --WVVSSMMLGAAIGALFNGWLSFRLGRKYSLMAGAILFVLGSIGSAFATSVEMLIAARV 120

Qy        129 LTGLACGVASLVAPVYISEIAYPAVRGLLGSCVQLMVVVGILLAYLAGWVLEW--RWLAV 186
              : |:| |:||  ||:|:||:|   ||| : |  |||| :||:||:|:     :   | |:
Db        121 VLGIAVGIASYTAPLYLSEMASENVRGKMISMYQLMVTLGIVLAFLSDTAFSYSGNWRAM 180
```

```
Qy          187 LG--CVPPSLMLLLMCFMPETPRFLLTQHRRQEAMAALRFLWGSEQGWEDPPIGAEQSFH 244
                ||   :|  |:::|: |:| :||:|  : |  ||   || |  : :   :      :|
Db          181 LGVLALPAVLLIILVVFLPNSPRWLAEKGRHIEAEEVLRMLRDTSEKAREELNEIRESLK 240

Qy          245 L-----ALLR-QPGIYKPFIIGVSLMAFQQLSGVNAVMFYAETIFEEAKF---KDSSLAS 295
                |     || :    : :   :|: | | || :|:| :|:||  ||: | |   :   :|:
Db          241 LKQGGWALFKINRNVRRAVFLGMLLQAMQQFTGMNIIMYYAPRIFKMAGFTTTEQQMIAT 300

Qy          296 VVVGVIQVLFTAVAALIMDRAGRRLLLVLSGVVMVFSTSAFGAYFKLTQGGPGNSSHVAI 355
                :|||:  :  | :|   :|:|||:  | :   ||   |   |  : | |   |
Db          301 LVVGLTFMFATFIAVFTVDKAGRKPALKIGFSVMALGTLVLG--YCLMQFDNGT------ 352

Qy          356 SAPVSAQPVDASVGLAWLAVGSMCLFIAGFAVGWGPIPWLLMSEIFPLHVKGVATGICVL 415
                          || ||:||:||   : |||:|:   |: |:| ||| ||  :
Db          353 ----------ASSGLSWLSVGMTMMCIAGYAMSAAPVVWILCSEIQPLKCRDFGITCSTT 402

Qy          416 TNWLMAFLVTKEFSSLMEVLRPYGAFWLASAFCIFSVLFTLFCVPETKGKTLEQI 470
                |||:   ::   | :|:: :   | ||| :|  |  |  | : :||||  ||| |
Db          403 TNWVSNMIIGATFLTLLDSIGAAGTFWLYTALNIAFVGITFWLIPETKNVTLEHI 457


RESULT 9
F65079
galactose-proton symport (galactose transporter) - Escherichia coli (strain K-12)
C;Species: Escherichia coli
C;Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
C;Accession: F65079
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; C
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617; PMID:9278503
A;Accession: F65079
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-464 <BLAT>
A;Cross-references: GB:AE000377; GB:U00096; NID:g2367178; PIDN:AAC75980.1; PID:g1789312;
A;Experimental source: strain K-12, substrain MG1655
C;Genetics:
A;Gene: galP
C;Superfamily: glucose transport protein

  Query Match             24.1%;  Score 591.5;  DB 2;  Length 464;
  Best Local Similarity   32.4%;  Pred. No. 2.9e-36;
  Matches  155;  Conservative   83;  Mismatches  175;  Indels   65;  Gaps   11;

Qy           27 FLAAFAAALGPLSFGFALGYSSPAIPSLQRAAPPAPRLDDAAASWFGAVVTLGAAAGGVL 86
                |   | |||  | ||  :|  : |:| :        ::      |  : :  ||| | |
Db           16 FFVCFLAALAGLLFGLDIGVIAGALPFIADEF----QITSHTQEWVVSSMMFGAAVGAVG 71

Qy           87 GGWLVDRAGRKLSLLLCSVPFVAGFAVITAAQDVWMLLGGRLLTGLACGVASLVAPVYIS 146
                 |||  : ||| ||:: :: ||||     || :| :|:  |:| ||| ||||  ||:|:|
Db           72 SGWLSFKLGRKKSLMIGAILFVAGSLFSAAAPNVEVLILSRVLLGLAVGVASYTAPLYLS 131

Qy          147 EIAYPAVRGLLGSCVQLMVVVGILLAYLA----GWVLEWRWLAVLGCVPPSLMLLLMCFM 202
                |||   :|| : |  |||: :||| |||:     :   |||:  :  :|  |:|: : |:
Db          132 EIAPEKIRGSMISMYQLMITIGILGAYLSDTAFSYTGAWRWMLGVIIIPAILLLIGVFFL 191

Qy          203 PETPRFLLTQHRRQEAMAAL------------------RFLWGSEQGWEDPPIGAEQSFH 244
                |::||:   : |  :|   |                    |   : ||
Db          192 PDSPRWFAAKRRFVDAERVLLRLRDTSAEAKRELDEIRESLQVKQSGW------------ 239

Qy          245 LALLRQPGIYKPFI-IGVSLMAFQQLSGVNAVMFYAETIFEEAKFKDSS---LASVVVGV 300
                 || ::   ::  : :|| |   || :|:| :|:||  ||| | : :::     :|:||:
Db          240 -ALFKENSNFRRAVFLGVLLQVMQQFTGMNVIMYYAPKIFELAGYTNTTEQMWGTVIVGL 298

Qy          301 IQVLFTAVAALIMDRAGRRLLLVLSGVVMVFSTSAFGAYFKLTQGGPGNSSHVAISAPVS 360
                  || | :|  ::|| ||:  | |  :||     | |       |  |   |: | :| |
Db          299 TNVLATFIAIGLVDRWGRKPTLTLGFLVM-----AAG------MGVLGTMMHIGIHSP-S 346

Qy          361 AQPVDASVGLAWLAVGSMCLFIAGFAVGWGPIPWLLMSEIFPLHVKGVATGICVLTNWLM 420
                ||         : |:  : :|| |||:  ||: |:| ||| ||  :         |||:
Db          347 AQ---------YFAIAMLLMFIVGFAMSAGPLIWVLCSEIQPLKGRDFGITCSTATNWIA 397

Qy          421 AFLVTKEFSSLMEVLRPYGAFWLASAFCIFSVLFTLFCVPETKGKTLEQITAH-FEGR 477
                  :|   | :::  |     ||: :|  :  :| ||: |||||  :|| |  :  :||
Db          398 NMIVGATFLTMLNTLGNANTFWVYAALNVLFILLTLWLVPETKHVSLEHIERNLMKGR 455


RESULT 10
C91106
galactose-proton symport of transport system ECs3819 [imported] - Escherichia coli (stra
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C;Accession: C91106
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: C91106
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-464 <HAY>
A;Cross-references: GB:BA000007; PIDN:BAB37242.1; PID:g13363291; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs3819
C;Superfamily: glucose transport protein

  Query Match             23.9%;  Score 587.5;  DB 2;  Length 464;
  Best Local Similarity   32.2%;  Pred. No. 5.6e-36;
  Matches  154;  Conservative   83;  Mismatches  176;  Indels   65;  Gaps   11;

Qy           27 FLAAFAAALGPLSFGFALGYSSPAIPSLQRAAPPAPRLDDAAASWFGAVVTLGAAAGGVL 86
                |   | |||  | ||  :|  : |:| :        ::      |  : :  ||| | |
Db           16 FFVCFLAALAGLLFGLDIGVIAGALPFIADEF----QITSHTQEWVVSSMMFGAAVGAVG 71

Qy           87 GGWLVDRAGRKLSLLLCSVPFVAGFAVITAAQDVWMLLGGRLLTGLACGVASLVAPVYIS 146
                 |||  : ||| ||:: :: ||||     || :| :|:  |:| ||| ||||  ||:|:|
Db           72 SGWLSFKLGRKKSLMIGAILFVAGSLFSAAAPNVEVLILSRVLLGLAVGVASYTAPLYLS 131

Qy          147 EIAYPAVRGLLGSCVQLMVVVGILLAYLA----GWVLEWRWLAVLGCVPPSLMLLLMCFM 202
                |||   :|| : |  |||: :||| |||:     :   |||:  :  :|  |:|: : |:
Db          132 EIAPEKIRGSMISMYQLMITIGILGAYLSDTAFSYTGAWRWMLGVIIIPAILLLIGVFFL 191

Qy          203 PETPRFLLTQHRRQEAMAAL------------------RFLWGSEQGWEDPPIGAEQSFH 244
                |::||:   : |  :|   |                    |   : ||
Db          192 PDSPRWFAAKRRFVDAERVLLRLRDTSAEAKRELDEIRESLQVKQSGW------------ 239

Qy          245 LALLRQPGIYKPFI-IGVSLMAFQQLSGVNAVMFYAETIFEEAKFKDSS---LASVVVGV 300
                 || ::   ::  : :|| |   || :|:| :|:||  ||| | : :::     :|:||:
Db          240 -ALFKENSNFRRAVFLGVLLQVMQQFTGMNVIMYYAPKIFELAGYTNTTEQMWGTVIVGL 298

Qy          301 IQVLFTAVAALIMDRAGRRLLLVLSGVVMVFSTSAFGAYFKLTQGGPGNSSHVAISAPVS 360
                  || | :|  ::|| ||:  | |  :||     | |       |  |   |: | :| |
Db          299 TNVLATFIAIGLVDRWGRKPTLTLGFLVM-----AAG------MGVLGTMMHIGIHSP-S 346

Qy          361 AQPVDASVGLAWLAVGSMCLFIAGFAVGWGPIPWLLMSEIFPLHVKGVATGICVLTNWLM 420
                ||         : |:  : :|| |||:  ||: |:| ||| ||  :         |||:
Db          347 AQ---------YFAIAMLLMFIVGFAMSAGPLIWVLCSEIQPLKGRDFGITCSTATNWIA 397

Qy          421 AFLVTKEFSSLMEVLRPYGAFWLASAFCIFSVLFTLFCVPETKGKTLEQITAH-FEGR 477
                  :|   | :::  |     ||: :   :  :| ||: |||||  :|| |  :  :||
Db          398 NMIVGATFLTMLNTLGNANTFWVYAGLNVLFILLTLWLVPETKHVSLEHIERNLMKGR 455


RESULT 11
F85951
galactose-proton symport of transport system [imported] - Escherichia coli (strain O157
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C;Accession: F85951
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhe
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca
Nature 409, 529-533, 2001
```

```
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: F85951
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-464 <STO>
A;Cross-references: GB:AE005174; NID:g12517486; PIDN:AAG58074.1; GSPDB:GN00145; UWGP:Z42
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: galP
C;Superfamily: glucose transport protein

  Query Match             23.9%;  Score 587.5;  DB 2;  Length 464;
  Best Local Similarity   32.2%;  Pred. No. 5.6e-36;
  Matches  154;  Conservative   83;  Mismatches  176;  Indels   65;  Gaps   11;

Qy         27 FLAAFAAALGPLSFGFALGYSSPAIPSLQRAAPPAPRLDDAAASWFGAVVTLGAAAGGVL 86
              |   | |||  | ||  :|  : |:| :        ::      |  : :  ||| | |
Db         16 FFVCFLAALAGLLFGLDIGVIAGALPFIADEF----QITSHTQEWVVSSMMFGAAVGAVG 71

Qy         87 GGWLVDRAGRKLSLLLCSVPFVAGFAVITAAQDVWMLLGGRLLTGLACGVASLVAPVYIS 146
               |||  : ||| ||:: :: ||||     || :| :|:  |:| ||| ||||  ||:|:|
Db         72 SGWLSFKLGRKKSLMIGAILFVAGSLFSAAAPNVEVLILSRVLLGLAVGVASYTAPLYLS 131

Qy        147 EIAYPAVRGLLGSCVQLMVVVGILLAYLA----GWVLEWRWLAVLGCVPPSLMLLLMCFM 202
              |||   :|| : |  |||: :||| |||:     :   |||:  :  :|  |:|: : |:
Db        132 EIAPEKIRGSMISMYQLMITIGILGAYLSDTAFSYTGAWRWMLGVIIIPAILLLIGVFFL 191

Qy        203 PETPRFLLTQHRRQEAMAAL--------------------RFLWGSEQGWEDPPIGAEQSFH 244
              |::||:   : |  :|   |                    |   :  ||
Db        192 PDSPRWFAAKRRFVDAERVLLRLRDTSAEAKRELDEIRESLQVKQSGW------------ 239

Qy        245 LALLRQPGIYKPFI-IGVSLMAFQQLSGVNAVMFYAETIFEEAKFKDSS---LASVVVGV 300
               || ::   ::  : :|| |   || :|:| :|:||  ||| | : :::     :|:||:
Db        240 -ALFKENSNFRRAVFLGVLLQVMQQFTGMNVIMYYAPKIFELAGYTNTTEQMWGTVIVGL 298

Qy        301 IQVLFTAVAALIMDRAGRRLLLVLSGVVMVFSTSAFGAYFKLTQGGPGNSSHVAISAPVS 360
                || | :|  ::|| ||:  | |  :||     | |       |  |   |: | :| |
Db        299 TNVLATFIAIGLVDRWGRKPTLTLGFLVM-----AAG------MGVLGTMMHIGIHSP-S 346

Qy        361 AQPVDASVGLAWLAVGSMCLFIAGFAVGWGPIPWLLMSEIFPLHVKGVATGICVLTNWLM 420
              ||         : |:  : :|| |||:  |:  |:| ||| ||  :         |||:
Db        347 AQ---------YFAIAMLLMFIVGFAMSAGPLIWVLCSEIQPLKGRDFGITCSTATNWIA 397

Qy        421 AFLVTKEFSSLMEVLRPYGAFWLASAFCIFSVLFTLFCVPETKGKTLEQITAH-FEGR 477
                :|   | :::  |     ||: :   :  :| ||: |||||  :|| |  :  :||
Db        398 NMIVGATFLTMLNTLGNANTFWVYAGLNVLFILLTLWLVPETKHVSLEHIERNLMKGR 455

RESULT 12
T52132
probable sugar transporter protein ERD6 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 20-Oct-2000 #sequence_revision 20-Oct-2000 #text_change 03-Nov-2000
C;Accession: T52132
R;Kiyosue, T.; Abe, H.; Yamaguchi-Shinozaki, K.; Shinozaki, K.
Biochim. Biophys. Acta 1370, 187-191, 1998
A;Title: ERD6, a cDNA clone for an early dehydration-induced gene of Arabidopsis, encode
A;Reference number: Z25973; MUID:98213606; PMID:9545564
A;Accession: T52132
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-496 <KIY>
A;Cross-references: EMBL:D89051; PIDN:BAA25989.1
C;Genetics:
A;Gene: ERD6
C;Superfamily: glucose transport protein
C;Keywords: transmembrane protein; transport protein

  Query Match             23.8%;  Score 584.5;  DB 2;  Length 496;
  Best Local Similarity   33.5%;  Pred. No. 1e-35;
  Matches  160;  Conservative   73;  Mismatches  169;  Indels   75;  Gaps   11;
```

```
Qy         26 VFLAAFAAALGPLSFGFALGYSSPAIPSLQRAAPPAPRLDDAAASWFGAVVTLGAAAGGV 85
              |||: | |  |    |  :|:|| |    |        |  |  | ||:::|||   | |
Db         58 VFLSTFVAVSGSFCTGCGVGFSSGA----QAGITKDLSLSVAEYSMFGSILTLGGLIGAV 113

Qy         86 LGGWLVDRAGRKLSLLLCSVPFVAGFAVITAAQDVWMLLGGRLLTGLACGVASLVAPVYI 145
                | : |  ||| ::| |    : |:  :  ||:   |  |||| |:  |: | | ||||
Db        114 FSGKVADVLGRKRTMLFCEFFCITGWLCVALAQNAMWLDCGRLLLGIGVGIFSYVIPVYI 173

Qy        146 SEIAYPAVRGLLGSCVQLMVVVGILLAYLAGWVLEWRWLAVLGCVPPSLMLLLMCFMPET 205
              :|||   |||      |||   || | :: |  : || | |:| ||    :  : |:||:
Db        174 AEIAPKHVRGSFVFANQLMQNCGISLFFIIGNFIPWRLLTVVGLVPCVFHVFCLFFIPES 233

Qy        206 PRFLLTQHRRQEAMAALRFLWGSEQGWEDPPIGAEQS--------------FHLALLRQP 251
              ||:|    | :|  ::|: | ||     |  |  | :                :: | |
Db        234 PRWLAKLGRDKECRSSLQRLRGS-----DVDISREANTIRDTIDMTENGGETKMSELFQR 288

Qy        252 GIYKPFIIGVSLMAFQQLSGVNAVMFYAETIFEEAKFKDSSLASVVVGVIQVLFTAVAAL 311
                  | |||| ||  ||| | : | :|| ::| :  |  |:: : |:  | |    :| :
Db        289 RYAYPLIIGVGLMFLQQLCGSSGVTYYASSLFNKGGF-PSAIGTSVIATIMVPKAMLATV 347

Qy        312 IMDRAGRRLLLVLSGVVMVFST---------SAFGAYFKLTQGGPGNSSHVAISAPVSAQ 362
              ::|: ||| ||: |   |  |           :||   :||
Db        348 LVDKMGRRTLLMASCSAMGLSALLLSVSYGFQSFGILPELT------------------- 388

Qy        363 PVDASVGLAWLAVGSMCLFIAGFAVGWGPIPWLLMSEIFPLHVKGVATGICVLTNWLMAF 422
              |:   :|:    :|     |  ||:| | :|::|:||||::||   |  :  :||||  :
Db        389 PIFTCIGV----LG----HIVSFAMGMGGLPWIIMAEIFPMNVKVSAGTLVTVTNWLFGW 440

Qy        423 LVTKEFSSLMEVLRPYGAFWLASA-FCIF------SVLFTLFCVPETKGKTLEQITA 472
              ::|  |: ::|        | ||  | ||      |::|  | ||||||::||:| |
Db        441 IITYTFNFMLE--------WNASGMFLIFSMVSASSIVFIYFLVPETKGRSLEEIQA 489

RESULT 13
T43400
myo-inositol transporter - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 20-Oct-2000
C;Accession: T43400; T39079
R;Niederberger, C.
submitted to the EMBL Data Library, June 1996
A;Reference number: Z22484
A;Accession: T43400
A;Molecule type: DNA
A;Residues: 1-575 <NIE>
A;Cross-references: EMBL:X98622; PIDN:CAA67211.1
A;Experimental source: strain 968 h90
R;Gentles, S.; Churcher, C.M.; Wood, V.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, September 1997
A;Reference number: Z21826
A;Accession: T39079
A;Molecule type: DNA
A;Residues: 290-575 <GEN>
A;Cross-references: EMBL:Z99532; PIDN:CAB16718.1; GSPDB:GN00066; SPDB:SPAC7D4.01
A;Experimental source: strain 972h-; cosmid c7D4
C;Genetics:
A;Gene: itr1; SPDB:SPAC7D4.01
A;Map position: 1
C;Superfamily: maltose transport protein MAL61
C;Keywords: sugar transport; transmembrane protein

  Query Match             23.8%;  Score 584.5;  DB 2;  Length 575;
  Best Local Similarity   29.8%;  Pred. No. 1.2e-35;
  Matches  147;  Conservative   90;  Mismatches  201;  Indels   55;  Gaps   11;

Qy          5 DPEETQPLLGPPGGSAPRGRRVFLAAFAAALGPLSFGFALGYSSPAIPSLQRAAPPAPRL 64
              |  | ||:           : |:: |||| :| | ||:  |  | |:  :  :      |
Db         76 DTVEDQPV----------SKWVWVLAFAAGIGGLLFGYDTGVISGALVVIGTSL-GGHEL 124

Qy         65 DDAAASWFGAVVTLGAAAGGVLGGWLVDRAGRKLSLLLCSVPFVAGFAVITAAQDVWMLL 124
               :    :  :  :|||  ||:: | | |  |||  : : |:  : |  |   |  :| ::
Db        125 TNGGKEFITSATSLGALLGGIIAGALADFFGRKPVIAIASIIIIVGSIVQVTAHHLWHMI 184
```

```
Qy        125 GGRLLTGLACGVASLVAPVYISEIAYPAVRGLLGSCVQLMVVVGILLAYLAGWVLE---- 180
               || : |   |:|||: |:|:||||   :|| |     |::  | ::||      |
Db        185 VGRFVIGWGVGIASLIIPLYLSEIAPSKIRGRLVIIYVLLITAGQVIAYGIDTAFEHVHN 244

Qy        181 -WRWLAVLGCVPPSLMLLLMCFMPETPRFLLTQHRRQEAMAALRFLWGS---------- 228
               |||:  |  || :  | :: ::||:|| |: : | |||   |  :: :
Db        245 GWRWMVGLAMVPAAFQLFILIWLPESPRLLVKKERSQEAYNTLARIYPTAHPYEIKTKLY 304

Qy        229 --EQGWEDPPIGAE-----QSFHLALLRQPGIYKPFIIGVSLMAFQQLSGVNAVMFYAET 281
                ::|  ||  |:      ::|   |   |  :: |:    | | ||||| |::|::: |
Db        305 LIQEGVRDPFSGSRWQKIVKTFK-ELYFNPSNFRALILACGLQAMQQLSGFNSLMYFSST 363

Qy        282 IFEEAKFKDSSLASVVVGVIQVLFTAVAALIMDRAGRRLLLVLS--GVVMVFSTSAFGAY 339
              |||   | : :   :::     :|| ||  ::|  |||:||:|:  |::      |   :
Db        364 IFEVVGFNNPTATGLIIAATNFVFTIVAFGVIDFFGRRILLLLTVWGMIAALIVCAVAFH 423

Qy        340 FKLTQGGPGNSSHVAISAPVSAQPVDASVGLAWLAVG--SMCLFIAGFAVGWGPIPWLLM 397
              | | :   || :              :    ||  |   || :::| :| | | :||
Db        424 F-LPKDENGNYT--------------SGQSNAWAIVVLISMIVYVASYASGLGNLPW-QQ 467

Qy        398 SEIFPLHVKGVATGICVLTNWLMAFLVTKEFSSLMEVLRPYGAFWLASAFCIFSVLFTLF 457
              ||:||: |:|: ||:    ||     :   | :||  : | | | |    |    |  ||
Db        468 SELFPMSVRGLGTGMSTAVNWAGNLGIGASFLTLMSEITPTGTFALYGGLCFLGWLGALF 527

Qy        458 CVPETKGKTLEQI 470
              | |:    |:|:|
Db        528 CYPDLTDYTIEEI 540


RESULT 14
AC0877
galactose-proton symport (galactose transporter) STY3244 [imported] - Salmonella enteric
C;Species: Salmonella enterica subsp. enterica serovar Typhi
A;Note: this species has also been called Salmonella typhi
C;Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C;Accession: AC0877
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
, S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A;Reference number: AB0502; MUID:21534947; PMID:11677608
A;Accession: AC0877
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-464 <PAR>
A;Cross-references: GB:AL513382; PIDN:CAD02915.1; PID:g16504168; GSPDB:GN00176
C;Genetics:
A;Gene: STY3244
C;Superfamily: glucose transport protein

  Query Match             23.7%;  Score 583.5;  DB 2;  Length 464;
  Best Local Similarity   31.6%;  Pred. No. 1.1e-35;
  Matches  151;  Conservative   84;  Mismatches  178;  Indels   65;  Gaps   10;

Qy         27 FLAAFAAALGPLSFGFALGYSSPAIPSLQRAAPPAPRLDDAAASWFGAVVTLGAAAGGVL 86
              |   | |||  | ||  :|  : |:| :        ::      |  : :  ||| | |
Db         16 FFVCFLAALAGLLFGLDIGVIAGALPFITDEF----QITAHTQEWVVSSMMFGAAVGAVG 71

Qy         87 GGWLVDRAGRKLSLLLCSVPFVAGFAVITAAQDVWMLLGGRLLTGLACGVASLVAPVYIS 146
               |||  : ||| ||:: :: ||||     || :|:|:  |:| ||| ||||  ||:|:|
Db         72 SGWLSFKLGRKKSLMIGAILFVAGSLFSAAAPNVEVLIISRVLLGLAVGVASYTAPLYLS 131

Qy        147 EIAYPAVRGLLGSCVQLMVVVGILLAYLAGWVLE----WRWLAVLGCVPPSLMLLLMCFM 202
              |||   :|| : |  |||: :||| |||:         |||:  :  :|  |:|: : |:
Db        132 EIAPEKIRGSMISMYQLMITIGILGAYLSDTAFSYSGAWRWMLGVIIIPAILLLIGVFFL 191

Qy        203 PETPRFLLTQHRRQEAMAAL------------------RFLWGSEQGWEDPPIGAEQSFH 244
              |::||:   :  |  :|   |                    |   : ||
Db        192 PDSPRWFAAKRRFHDAERVLLRLRDTSAEAKRELDEIRESLQVKQSGW------------ 239

Qy        245 LALLRQPGIYKPFI-IGVSLMAFQQLSGVNAVMFYAETIFEEAKFKDSS---LASVVVGV 300
```

```
               || ::   ::  : :|: |   || :|:| :|:||  ||| | : :::     :|:||:
Db        240 -ALFKENSNFRRAVFLGILLQVMQQFTGMNVIMYYAPKIFELAGYTNTTEQMWGTVIVGL 298

Qy        301 IQVLFTAVAALIMDRAGRRLLLVLSGVVMVFSTSAFGAYFKLTQGGPGNSSHVAISAPVS 360
                || | :|  ::|| ||:  | |  :||            :  |  |   |: | :| |
Db        299 TNVLATFIAIGLVDRWGRKPTLTLGFLVMA-----------IGMGILGTMMHIGIHSP-S 346

Qy        361 AQPVDASVGLAWLAVGSMCLFIAGFAVGWGPIPWLLMSEIFPLHVKGVATGICVLTNWLM 420
              ||         : |:  : :|| |||:  ||:|:|  ||| ||  :         |||:
Db        347 AQ---------YFAIAMLLMFIIGFAMSAGPLIWVLCSEIQPLKGRDFGITCSTATNWIA 397

Qy        421 AFLVTKEFSSLMEVLRPYGAFWLASAFCIFSVLFTLFCVPETKGKTLEQITAH-FEGR 477
                :|   | :::  |     ||: :   :  :|  ||: |||||  :|| |  :  :||
Db        398 NMIVGATFLTMLNNLGNANTFWVYAGLNVLFILLTLWLVPETKHVSLEHIERNLMKGR 455


RESULT 15
G84864
probable membrane transporter [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C;Accession: G84864
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: G84864
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-521 <STO>
A;Cross-references: GB:AE002093; NID:g2289003; PIDN:AAB64332.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g43330
A;Map position: 2
C;Superfamily: glucose transport protein

  Query Match             23.7%;  Score 582;  DB 2;  Length 521;
  Best Local Similarity   31.0%;  Pred. No. 1.6e-35;
  Matches  147;  Conservative   76;  Mismatches  191;  Indels   60;  Gaps    9;

Qy         33 AALGPLSFGFALGYSSPAIPSL--------QRAAPPAPRLDDAAASWFGAVVTL---GAA 81
              | :| | ||:  |  | |:  :        | :      :    :|    :|::   ||
Db         38 AGIGGLLFGYDTGVISGALLYIKDDFEVVKQSSFLQVYNVSSFTSSKLETIVSMALVGAM 97

Qy         82 AGGVLGGWLVDRAGRKLSLLLCSVPFVAGFAVITAAQDVWMLLGGRLLTGLACGVASLVA 141
               |   |||: |  ||| : |   | | ||  |: || | ::|: |||| ||  ||||: |
Db         98 IGAAAGGWINDYYGRKKATLFADVVFAAGAIVMAAAPDPYVLISGRLLVGLGVGVASVTA 157

Qy        142 PVYISEIAYPAVRGLLGSCVQLMVVVGILLAYLAGWVL-----EWRWLAVLGCVPPSLML 196
              ||||:| :   ||| | |   ||:  |  |:||           |||:  :  ||  :
Db        158 PVYIAEASPSEVRGGLVSTNVLMITGGQFLSYLVNSAFTQVPGTWRWMLGVSGVPAVIQF 217

Qy        197 LLMCFMPETPRFLLTQHRRQEAMAALRFLWG---------------SEQGWEDPPIGAEQ 241
              :|| ||||:||:|  ::|: ||:  |   :                 |:      :|
Db        218 ILMLFMPESPRWLFMKNRKAEAIQVLARTYDISRLEDEIDHLSAAEEEEKQRKRTVG--- 274

Qy        242 SFHLALLRQPGIYKPFIIGVSLMAFQQLSGVNAVMFYAETIFEEAKFKDSSLA---SVVV 298
                :| :  |   :   |: |  | |||| :|:| ||:|: || : | |  : ||  |:: |
Db        275 --YLDVFRSKELRLAFLAGAGLQAFQQFTGINTVMYYSPTIVQMAGFHSNQLALFLSLIV 332

Qy        299 GVIQVLFTAVAALIMDRAGRRLLLVLSGVVMVFSTSAFGAYFKLTQGGPGNSSHVAISAP 358
                :    | |    :|  ||: | :         :| ||    |          : :|
Db        333 AAMNAAGTVVGIYFIDHCGRKKLAL---------SSLFGVIISL----------LILSVS 373

Qy        359 VSAQPVDASVG--LAWLAVGSMCLFIAGFAVGWGPIPWLLMSEIFPLHVKGVATGICVLT 416
                 |   :| |    ||||  : |:|  || | ||:|| : |||:|   :|:  |:
Db        374 FFKQSETSSDGGLYGWLAVLGLALYIVFFAPGMGPVPWTVNSEIYPQQYRGICGGMSATV 433

Qy        417 NWLMAFLVTKEFSSLMEVLRPYGAFWLASAFCIFSVLFTLFCVPETKGKTLEQI 470
              ||:   :| : | :: |       | : :   : :|:| :  ||||:| |  ::
Db        434 NWISNLIVAQTFLTIAEAAGTGMTFLILAGIAVLAVIFVIVFVPETQGLTFSEV 487
```

```
Search completed: September 27, 2004, 19:30:11
Job time : 51 secs
```

```
                    GenCore version 5.1.6
           Copyright (c) 1993 - 2004  Compugen Ltd.

OM protein - protein search, using sw model

Run on:         September 27, 2004, 19:09:21 ; Search time 25 Seconds
                                             (without alignments)
                                             993.499 Million cell updates/sec

Title:          US-09-886-954A-1
Perfect score:  2457
Sequence:       1 MTPEDPEETQPLLGPPGGSA.........CVPETKGKTLEQITAHFEGR 477

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters:      141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       SwissProt_42:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                    SUMMARIES
                   %
Result           Query
  No.    Score   Match Length DB  ID                      Description
------------------------------------------------------------------------------
    1     2445    99.5    477  1  GTR8_HUMAN              Q9ny64 homo sapien
    2   2162.5    88.0    478  1  GTR8_RAT                Q9jjz1 rattus norv
    3     2156    87.7    477  1  GTR8_MOUSE              Q9jif3 mus musculu
    4   1537.5    62.6    334  1  GTR8_BOVIN              P58354 bos taurus
    5      948    38.6    507  1  GTR6_HUMAN              Q9ugq3 homo sapien
    6      607    24.7    461  1  CSBC_BACSU              P46333 bacillus su
    7    606.5    24.7    472  1  ARAE_KLEOX              P45598 klebsiella
    8      603    24.5    472  1  ARAE_ECOLI              P09830 escherichia
    9    591.5    24.1    464  1  GALP_ECOLI              P37021 escherichia
   10    584.5    23.8    575  1  ITR1_SCHPO              Q10286 schizosacch
   11    579.5    23.6    494  1  GTR3_BOVIN              P58352 bos taurus
   12    575.5    23.4    494  1  GTR3_SHEEP              P47843 ovis aries
   13    574.5    23.4    493  1  GTR3_RAT                Q07647 rattus norv
   14    570.5    23.2    496  1  GTR3_HUMAN              P11169 homo sapien
   15    569.5    23.2    493  1  GTR3_MOUSE              P32037 mus musculu
   16    565.5    23.0    495  1  GTR3_CANFA              P47842 canis famil
   17    554.5    22.6    464  1  ARAE_BACSU              P96710 bacillus su
   18      550    22.4    629  1  MYCT_HUMAN              Q96qe2 homo sapien
   19      546    22.2    457  1  XYLT_LACBR              O52733 lactobacill
   20      546    22.2    491  1  XYLE_ECOLI              P09098 escherichia
   21      543    22.1    523  1  GTR2_MOUSE              P14246 mus musculu
   22    542.5    22.1    522  1  GTR2_RAT                P12336 rattus norv
   23      542    22.1    534  1  HUP3_CHLKE              Q39525 chlorella k
   24      536    21.8    509  1  GTR4_HUMAN              P14672 homo sapien
   25    535.5    21.8    522  1  STP1_ARATH              P23586 arabidopsis
   26      531    21.6    523  1  STC_RICCO               Q41144 ricinus com
   27    530.5    21.6    468  1  GLCP_SYNY3              P15729 synechocyst
   28      530    21.6    509  1  GTR4_RAT                P19357 rattus norv
   29    529.5    21.6    540  1  HUP2_CHLKE              Q39524 chlorella k
   30      529    21.5    509  1  GTR4_MOUSE              P14142 mus musculu
   31    528.5    21.5    534  1  HUP1_CHLKE              P15686 chlorella k
   32    526.5    21.4    496  1  GTR3_CHICK              P28568 gallus gall
   33      523    21.3    612  1  ITR2_YEAST              P30606 saccharomyc
   34    520.5    21.2    492  1  GTR1_MOUSE              P17809 mus musculu
   35      516    21.0    584  1  ITR1_YEAST              P30605 saccharomyc
   36    515.5    21.0    492  1  GTR1_BOVIN              P27674 bos taurus
   37    514.5    20.9    492  1  GTR1_HUMAN              P11166 homo sapien
   38    513.5    20.9    492  1  GTR1_RAT                P11167 rattus norv
   39      513    20.9    524  1  GTR2_HUMAN              P11168 homo sapien
   40    512.5    20.9    490  1  GTR1_CHICK              P46896 gallus gall
   41    510.5    20.8    557  1  ITR2_SCHPO              P87110 schizosacch
   42    507.5    20.7    492  1  GTR1_RABIT              P13355 oryctolagus
   43      507    20.6    400  1  GTR3_RABIT              Q9xsc2 oryctolagus
   44    499.5    20.3    473  1  GLF_ZYMMO               P21906 zymomonas m
   45      498    20.3    533  1  GTR2_CHICK              Q90592 gallus gall

                                  ALIGNMENTS

RESULT 1
GTR8_HUMAN
ID   GTR8_HUMAN     STANDARD;      PRT;   477 AA.
AC   Q9NY64; Q9NSC4;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Solute carrier family 2, facilitated glucose transporter, member 8
DE   (Glucose transporter type 8) (Glucose transporter type X1).
GN   SLC2A8 OR GLUT8 OR GLUTX1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Testis;
RX   MEDLINE=20283667; PubMed=10821868;
RA   Doege H., Schuermann A., Bahrenberg G., Brauers A., Joost H.-G.;
RT   "GLUT8, a novel member of the sugar transport facilitator family with
RT   glucose transport activity.";
RL   J. Biol. Chem. 275:16275-16280(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20138191; PubMed=10671487;
RA   Ibberson M.R., Uldry M.A., Thorens B.;
RT   "GLUTX1, a novel mammalian glucose transporter expressed in the
RT   central nervous system and insulin-sensitive tissues.";
RL   J. Biol. Chem. 275:4607-4612(2000).
CC   -!- FUNCTION: Insulin-regulated facilitative glucose transporter.
CC       Binds cytochalasin B in a glucose-inhibitable manner. Seems to be
CC       a dual-specific sugar transporter as it is inhibitable by
CC       fructose (By similarity).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Principally
CC       intracellular. May move between intracellular vesicles and the
CC       plasma membrane. The dileucine internalization motif is critical
CC       for intracellular sequestration (By similarity).
CC   -!- TISSUE SPECIFICITY: Highly expressed in testis, but not in
CC       testicular carcinoma. Lower amounts present in most other tissues.
CC   -!- INDUCTION: In testis, downregulated by estrogen.
CC   -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY. GLUCOSE
CC       TRANSPORTERS SUBFAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Y17801; CAB89809.1; -.
DR   EMBL; AJ245937; CAB75702.1; -.
DR   Genew; HGNC:13812; SLC2A8.
DR   MIM; 605245; -.
DR   GO; GO:0005887; C:integral to plasma membrane; TAS.
```

```
DR   GO; GO:0005355; F:glucose transporter activity; TAS.
DR   GO; GO:0005975; P:carbohydrate metabolism; TAS.
DR   GO; GO:0015758; P:glucose transport; TAS.
DR   InterPro; IPR007114; MFS.
DR   InterPro; IPR005828; Sub_transporter.
DR   InterPro; IPR005829; Sug_transporter.
DR   InterPro; IPR003663; Sugar_transpt.
DR   Pfam; PF00083; sugar_tr; 1.
DR   PRINTS; PR00171; SUGRTRNSPORT.
DR   TIGRFAMs; TIGR00879; SP; 1.
DR   PROSITE; PS50850; MFS; 1.
DR   PROSITE; PS00216; SUGAR_TRANSPORT_1; 2.
DR   PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.
KW   Transport; Sugar transport; Transmembrane; Glycoprotein;
KW   Multigene family.
FT   DOMAIN        1     25       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     26     46       1 (POTENTIAL).
FT   DOMAIN       47     70       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     71     91       2 (POTENTIAL).
FT   DOMAIN       92     96       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     97    117       3 (POTENTIAL).
FT   DOMAIN      118    127       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    128    148       4 (POTENTIAL).
FT   DOMAIN      149    156       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    157    177       5 (POTENTIAL).
FT   DOMAIN      178    182       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    183    203       6 (POTENTIAL).
FT   DOMAIN      204    256       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    257    277       7 (POTENTIAL).
FT   DOMAIN      278    292       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    293    313       8 (POTENTIAL).
FT   DOMAIN      314    319       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    320    340       9 (POTENTIAL).
FT   DOMAIN      341    367       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    368    388       10 (POTENTIAL).
FT   DOMAIN      389    404       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    405    425       11 (POTENTIAL).
FT   DOMAIN      426    438       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    439    459       12 (POTENTIAL).
FT   DOMAIN      460    477       CYTOPLASMIC (POTENTIAL).
FT   SITE         12     13       DILEUCINE INTERNALIZATION MOTIF (BY
FT                                SIMILARITY).
FT   CARBOHYD    349    349       N-LINKED (GLCNAC. . .) (By similarity).
FT   CONFLICT    377    377       S -> N (IN REF. 2).
FT   CONFLICT    456    457       FS -> LF (IN REF. 2).
FT   CONFLICT    462    462       T -> I (IN REF. 2).
SQ   SEQUENCE   477 AA;  50792 MW;  0B480F94B40AEE76 CRC64;

  Query Match             99.5%;  Score 2445;  DB 1;  Length 477;
  Best Local Similarity   99.6%;  Pred. No. 1.6e-164;
  Matches  475;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 MTPEDPEETQPLLGPPGGSAPRGRRVFLAAFAAALGPLSFGFALGYSSPAIPSLQRAAPP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MTPEDPEETQPLLGPPGGSAPRGRRVFLAAFAAALGPLSFGFALGYSSPAIPSLQRAAPP 60

Qy         61 APRLDDAAASWFGAVVTLGAAAGGVLGGWLVDRAGRKLSLLLCSVPFVAGFAVITAAQDV 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 APRLDDAAASWFGAVVTLGAAAGGVLGGWLVDRAGRKLSLLLCSVPFVAGFAVITAAQDV 120

Qy        121 WMLLGGRLLTGLACGVASLVAPVYISEIAYPAVRGLLGSCVQLMVVVGILLAYLAGWVLE 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 WMLLGGRLLTGLACGVASLVAPVYISEIAYPAVRGLLGSCVQLMVVVGILLAYLAGWVLE 180

Qy        181 WRWLAVLGCVPPSLMLLLMCFMPETPRFLLTQHRRQEAMAALRFLWGSEQGWEDPPIGAE 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 WRWLAVLGCVPPSLMLLLMCFMPETPRFLLTQHRRQEAMAALRFLWGSEQGWEDPPIGAE 240

Qy        241 QSFHLALLRQPGIYKPFIIGVSLMAFQQLSGVNAVMFYAETIFEEAKFKDSSLASVVVGV 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QSFHLALLRQPGIYKPFIIGVSLMAFQQLSGVNAVMFYAETIFEEAKFKDSSLASVVVGV 300

Qy        301 IQVLFTAVAALIMDRAGRRLLLVLSGVVMVFSTSAFGAYFKLTQGGPGNSSHVAISAPVS 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 IQVLFTAVAALIMDRAGRRLLLVLSGVVMVFSTSAFGAYFKLTQGGPGNSSHVAISAPVS 360

Qy        361 AQPVDASVGLAWLAVGSMCLFIAGFAVGWGPIPWLLMSEIFPLHVKGVATGICVLTNWLM 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AQPVDASVGLAWLAVGSMCLFIAGFAVGWGPIPWLLMSEIFPLHVKGVATGICVLTNWLM 420

Qy        421 AFLVTKEFSSLMEVLRPYGAFWLASAFCIFSVLFTLFCVPETKGKTLEQITAHFEGR 477
              |||||||||||||||||||||||||||||||||||||  ||||||||||||||||||
Db        421 AFLVTKEFSSLMEVLRPYGAFWLASAFCIFSVLFTFSCVPETKGKTLEQITAHFEGR 477


RESULT 2
GTR8_RAT
ID   GTR8_RAT       STANDARD;      PRT;   478 AA.
AC   Q9JJZ1; Q9JMA6;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Solute carrier family 2, facilitated glucose transporter, member 8
DE   (Glucose transporter type 8) (Glucose transporter type X1).
GN   SLC2A8 OR GLUT8 OR GLUTX1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A., CHARACTERIZATION, AND MUTAGENESIS OF
RP   12-LEU-LEU-13.
RC   TISSUE=Testis;
RX   MEDLINE=20138191; PubMed=10671487;
RA   Ibberson M.R., Uldry M.A., Thorens B.;
RT   "GLUTX1, a novel mammalian glucose transporter expressed in the
RT   central nervous system and insulin-sensitive tissues.";
RL   J. Biol. Chem. 275:4607-4612(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Testis;
RA   Ishibashi K.;
RT   "Molecular cloning of a new putative glucose transporter.";
RL   Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   TISSUE SPECIFICITY.
RX   MEDLINE=20283667; PubMed=10821868;
RA   Doege H., Schuermann A., Bahrenberg C., Brauers A., Joost H.-G.;
RT   "GLUT8, a novel member of the sugar transport facilitator family with
RT   glucose transport activity.";
RL   J. Biol. Chem. 275:16275-16280(2000).
CC   -!- FUNCTION: Insulin-regulated facilitative glucose transporter.
CC       Binds cytochalasin B in a glucose-inhibitable manner. Seems to be
CC       a dual-specific sugar transporter as it is inhibitable by
CC       fructose.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Principally
CC       intracellular. May move between intracellular vesicles and the
CC       plasma membrane. The dileucine internalization motif is critical
CC       for intracellular sequestration.
CC   -!- TISSUE SPECIFICITY: Highly expressed in adult and pubertal testis,
CC       but not prepubertal testis. Moderate expression in hypothalamus,
CC       cerebellum, brainstem, hippocampus, and adrenal gland. Lower
CC       amounts present in most other tissues.
CC   -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY. GLUCOSE
CC       TRANSPORTERS SUBFAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AJ245935; CAB75729.1; -.
DR   EMBL; AB033418; BAA94383.1; -.
```

```
DR   InterPro; IPR007114; MFS.
DR   InterPro; IPR005828; Sub_transporter.
DR   InterPro; IPR005829; Sug_transporter.
DR   InterPro; IPR003663; Sugar_transpt.
DR   Pfam; PF00083; sugar_tr; 1.
DR   PRINTS; PR00171; SUGRTRNSPORT.
DR   TIGRFAMs; TIGR00879; SP; 1.
DR   PROSITE; PS50850; MFS; 1.
DR   PROSITE; PS00216; SUGAR_TRANSPORT_1; 2.
DR   PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.
KW   Transport; Sugar transport; Transmembrane; Glycoprotein;
KW   Multigene family.
FT   DOMAIN        1     25       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     26     46       1 (POTENTIAL).
FT   DOMAIN       47     70       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     71     91       2 (POTENTIAL).
FT   DOMAIN       92     96       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     97    117       3 (POTENTIAL).
FT   DOMAIN      118    127       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    128    148       4 (POTENTIAL).
FT   DOMAIN      149    156       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    157    177       5 (POTENTIAL).
FT   DOMAIN      178    182       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    183    203       6 (POTENTIAL).
FT   DOMAIN      204    257       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    258    278       7 (POTENTIAL).
FT   DOMAIN      279    293       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    294    314       8 (POTENTIAL).
FT   DOMAIN      315    320       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    321    341       9 (POTENTIAL).
FT   DOMAIN      342    368       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    369    389       10 (POTENTIAL).
FT   DOMAIN      390    405       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    406    426       11 (POTENTIAL).
FT   DOMAIN      427    439       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    440    460       12 (POTENTIAL).
FT   DOMAIN      461    478       CYTOPLASMIC (POTENTIAL).
FT   SITE         12     13       DILEUCINE INTERNALIZATION MOTIF.
FT   CARBOHYD    350    350       N-LINKED (GLCNAC. . .) (PROBABLE).
FT   MUTAGEN      12     13       LL->AA: CHANGES SUBCELLULAR LOCATION
FT                                MAINLY TO THE PLASMA MEMBRANE, THREREBY
FT                                INCREASING TRANSPORT ACTIVITY.
FT   CONFLICT     83     84       GG -> QGA (IN REF. 2).
SQ   SEQUENCE   478 AA;  51458 MW;  95841FC1F18C9EE9 CRC64;

  Query Match             88.0%;  Score 2162.5;  DB 1;  Length 478;
  Best Local Similarity   85.4%;  Pred. No. 1.1e-144;
  Matches  408;  Conservative   33;  Mismatches   36;  Indels    1;  Gaps    1;

Qy          1 MTPEDPEETQPLLGPPGGSAPRGRRVFLAAFAAALGPLSFGFALGYSSPAIPSLQRAAPP 60
              |:||||:||||||  ||  || ||||||| ||||||||||||||||||||||||:| |||
Db          1 MSPEDPQETQPLLRSPGARAPGGRRVFLATFAAALGPLSFGFALGYSSPAIPSLRRTAPP 60

Qy         61 APRLDDAAASWFGAVVTLGAAAGGVLGGWLVDRAGRKLSLLLCSVPFVAGFAVITAAQDV 120
              | || | |||||||||||||||||||||||:||||||||||||:|||| ||||||||:||
Db         61 ALRLGDTAASWFGAVVTLGAAAGGVLGGWLLDRAGRKLSLLLCTVPFVTGFAVITAARDV 120

Qy        121 WMLLGGRLLTGLACGVASLVAPVYISEIAYPAVRGLLGSCVQLMVVVGILLAYLAGWVLE 180
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||:||||||
Db        121 WMLLGGRLLTGLACGVASLVAPVYISEIAYPAVRGLLGSCVQLMVVTGILLAYVAGWVLE 180

Qy        181 WRWLAVLGCVPPSLMLLLMCFMPETPRFLLTQHRRQEAMAALRFLWGSEQGWEDPPIGAE 240
              ||||||||||||:|||||||:||||||||||||: ||||||||||||||:|||:||:|||
Db        181 WRWLAVLGCVPPTLMLLLMCYMPETPRFLLTQHQYQEAMAALRFLWGSEEGWEEPPVGAE 240

Qy        241 -QSFHLALLRQPGIYKPFIIGVSLMAFQQLSGVNAVMFYAETIFEEAKFKDSSLASVVVG 299
               | | ||:||:||::|| |||: || |||||||||:|||| |||||||||||||||| ||
Db        241 HQGFQLAMLRRPGVHKPLIIGICLMVFQQLSGVNAIMFYANTIFEEAKFKDSSLASVTVG 300

Qy        300 VIQVLFTAVAALIMDRAGRRLLLVLSGVVMVFSTSAFGAYFKLTQGGPGNSSHVAISAPV 359
              :||||||||||||||||||:||| ||||:|||| |||| |||||| || ||||| :  |:
Db        301 IIQVLFTAVAALIMDRAGRKLLLALSGVIMVFSMSAFGTYFKLTQSGPSNSSHVGLLVPI 360

Qy        360 SAQPVDASVGLAWLAVGSMCLFIAGFAVGWGPIPWLLMSEIFPLHVKGVATGICVLTNWL 419
              ||:| |   :||||||||||||||||||||||||||||||||||||:||||||:||||||
Db        361 SAEPADVHLGLAWLAVGSMCLFIAGFAVGWGPIPWLLMSEIFPLHIKGVATGVCVLTNWF 420

Qy        420 MAFLVTKEFSSLMEVLRPYGAFWLASAFCIFSVLFTLFCVPETKGKTLEQITAHFEGR 477
              |||||||||:|:||:||||||||| :|||| ||||||  ||||||:||||||||||||
Db        421 MAFLVTKEFNSIMEILRPYGAFWLTAAFCILSVLFTLTFVPETKGRTLEQITAHFEGR 478


RESULT 3
GTR8_MOUSE
ID   GTR8_MOUSE     STANDARD;      PRT;   477 AA.
AC   Q9JIF3; Q9JJP4; Q9JJZ0;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Solute carrier family 2, facilitated glucose transporter, member 8
DE   (Glucose transporter type 8) (Glucose transporter type X1).
GN   SLC2A8 OR GLUT8 OR GLUTX1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20138191; PubMed=10671487;
RA   Ibberson M.R., Uldry M.A., Thorens B.;
RT   "GLUTX1, a novel mammalian glucose transporter expressed in the
RT   central nervous system and insulin-sensitive tissues.";
RL   J. Biol. Chem. 275:4607-4612(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Testis;
RX   MEDLINE=20283667; PubMed=10821868;
RA   Doege H., Schuermann A., Bahrenberg C., Brauers A., Joost H.-G.;
RT   "GLUT8, a novel member of the sugar transport facilitator family with
RT   glucose transport activity.";
RL   J. Biol. Chem. 275:16275-16280(2000).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129; TISSUE=Embryonic carcinoma;
RX   MEDLINE=20319023; PubMed=10860996;
RA   Carayannopoulos M.O., Chi M.M.-Y., Cui Y., Pingsterhaus J.M.,
RA   McKnight R.A., Mueckler M., Devaskar S.U., Moley K.H.;
RT   "GLUT8 is a glucose transporter responsible for insulin-stimulated
RT   glucose uptake in the blastocyst.";
RL   Proc. Natl. Acad. Sci. U.S.A. 97:7313-7318(2000).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129/Ola; TISSUE=Spleen;
RX   MEDLINE=21547794; PubMed=11689004;
RA   Scheepers A., Doege H., Joost H.-G., Schuermann A.;
RT   "Mouse GLUT8: genomic organization and regulation of expression in
RT   3T3-L1 adipocytes by glucose.";
RL   Biochem. Biophys. Res. Commun. 288:969-974(2001).
CC   -!- FUNCTION: Insulin-regulated facilitative glucose transporter.
CC       Binds cytochalasin B in a glucose-inhibitable manner. Seems to be
CC       a dual-specific sugar transporter as it is inhibitable by
CC       fructose.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Principally
CC       intracellular. May move between intracellular vesicles and the
CC       plasma membrane. The dileucine internalization motif is critical
CC       for intracellular sequestration (By similarity). Insulin induces a
CC       change in the intracellular localization and gives rise to
CC       insertion in the plasma membrane.
CC   -!- TISSUE SPECIFICITY: Highest level of expression in placenta and
CC       testis. Highly expressed in adult and pubertal testis, but not
CC       prepubertal testis. Lower levels of expression in brain, liver,
CC       heart, kidney, fat and skeletal muscle.
CC   -!- DEVELOPMENTAL STAGE: High expression in blastocysts.
CC   -!- INDUCTION: Inhibited under glucose deprivation.
CC   -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY. GLUCOSE
CC       TRANSPORTERS SUBFAMILY.
```

```
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; AJ245936; CAB75719.1; -.
DR   EMBL; Y17802; CAB89815.1; -.
DR   EMBL; AF232061; AAF78366.1; -.
DR   EMBL; AJ413951; CAC88690.1; -.
DR   MGD; MGI:1860103; Slc2a8.
DR   GO; GO:0005887; C:integral to plasma membrane; IDA.
DR   GO; GO:0005536; F:glucose binding; IDA.
DR   GO; GO:0005355; F:glucose transporter activity; IDA.
DR   GO; GO:0015758; P:glucose transport; IDA.
DR   GO; GO:0008286; P:insulin receptor signaling pathway; IDA.
DR   GO; GO:0001666; P:response to hypoxia; IDA.
DR   InterPro; IPR007114; MFS.
DR   InterPro; IPR005828; Sub_transporter.
DR   InterPro; IPR005829; Sug_transporter.
DR   InterPro; IPR003663; Sugar_transpt.
DR   Pfam; PF00083; sugar_tr; 1.
DR   PRINTS; PR00171; SUGRTRNSPORT.
DR   TIGRFAMs; TIGR00879; SP; 1.
DR   PROSITE; PS50850; MFS; 1.
DR   PROSITE; PS00216; SUGAR_TRANSPORT_1; 1.
DR   PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.
KW   Transport; Sugar transport; Transmembrane; Glycoprotein;
KW   Multigene family.
FT   DOMAIN        1     25       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     26     46       1 (POTENTIAL).
FT   DOMAIN       47     70       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     71     91       2 (POTENTIAL).
FT   DOMAIN       92     96       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     97    117       3 (POTENTIAL).
FT   DOMAIN      118    127       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    128    148       4 (POTENTIAL).
FT   DOMAIN      149    156       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    157    177       5 (POTENTIAL).
FT   DOMAIN      178    182       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    183    203       6 (POTENTIAL).
FT   DOMAIN      204    257       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    258    278       7 (POTENTIAL).
FT   DOMAIN      279    293       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    294    314       8 (POTENTIAL).
FT   DOMAIN      315    320       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    321    341       9 (POTENTIAL).
FT   DOMAIN      342    367       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    368    388       10 (POTENTIAL).
FT   DOMAIN      389    404       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    405    425       11 (POTENTIAL).
FT   DOMAIN      426    438       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    439    459       12 (POTENTIAL).
FT   DOMAIN      460    477       CYTOPLASMIC (POTENTIAL).
FT   SITE         12     13       DILEUCINE INTERNALIZATION MOTIF (BY
FT                                SIMILARITY).
FT   CARBOHYD    350    350       N-LINKED (GLCNAC. . .) (By similarity).
FT   CONFLICT     39     39       S -> N (IN REF. 1).
FT   CONFLICT     94     94       S -> A (IN REF. 2 AND 4).
FT   CONFLICT    429    429       S -> N (IN REF. 1).
SQ   SEQUENCE   477 AA;  51523 MW;  A3753FB34E452F9A CRC64;

  Query Match             87.7%;  Score 2156;  DB 1;  Length 477;
  Best Local Similarity   86.0%;  Pred. No. 3e-144;
  Matches  411;  Conservative   31;  Mismatches   34;  Indels    2;  Gaps    2;

Qy          1 MTPEDPEETQPLLGPPGGSAPRGRRVFLAAFAAALGPLSFGFALGYSSPAIPSLQRAAPP 60
              |:||||:|||||| ||    |||||||||:||||||||||||||||||||||||:| |||
Db          1 MSPEDPQETQPLLRPPEARTPRGRRVFLASFAAALGPLSFGFALGYSSPAIPSLRRTAPP 60
```

```
Qy         61 APRLDDAAASWFGAVVTLGAAAGGVLGGWLVDRAGRKLSLLLCSVPFVAGFAVITAAQDV 120
              | || | |||||||||||||||||:|||||:||:|||||||||:|||| ||||||||:||
Db         61 ALRLGDNAASWFGAVVTLGAAAGGILGGWLLDRSGRKLSLLLCTVPFVTGFAVITAARDV 120

Qy        121 WMLLGGRLLTGLACGVASLVAPVYISEIAYPAVRGLLGSCVQLMVVVGILLAYLAGWVLE 180
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||:||||||
Db        121 WMLLGGRLLTGLACGVASLVAPVYISEIAYPAVRGLLGSCVQLMVVTGILLAYVAGWVLE 180

Qy        181 WRWLAVLGCVPPSLMLLLMCFMPETPRFLLTQHRRQEAMAALRFLWGSEQGWEDPPIGAE 240
              ||||||||||||:|||||||:||||||||||||: ||||||||||||||:|||:||:|||
Db        181 WRWLAVLGCVPPTLMLLLMCYMPETPRFLLTQHQYQEAMAALRFLWGSEEGWEEPPVGAE 240

Qy        241 -QSFHLALLRQPGIYKPFIIGVSLMAFQQLSGVNAVMFYAETIFEEAKFKDSSLASVVVG 299
               | | |||||:|||||| |||:||| |||||||||:|||| :||||||||||||||| ||
Db        241 HQGFQLALLRRPGIYKPLIIGISLMVFQQLSGVNAIMFYANSIFEEAKFKDSSLASVTVG 300

Qy        300 VIQVLFTAVAALIMDRAGRRLLLVLSGVVMVFSTSAFGAYFKLTQGGPGNSSHVAISAPV 359
              :|||||||||||||||||||||| ||||:|||| |||| ||||||  | ||||| :  |:
Db        301 IIQVLFTAVAALIMDRAGRRLLLALSGVIMVFSMSAFGTYFKLTQSLPSNSSHVGL-VPI 359

Qy        360 SAQPVDASVGLAWLAVGSMCLFIAGFAVGWGPIPWLLMSEIFPLHVKGVATGICVLTNWL 419
              :|:|||  |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        360 AAEPVDVQVGLAWLAVGSMCLFIAGFAVGWGPIPWLLMSEIFPLHVKGVATGICVLTNWF 419

Qy        420 MAFLVTKEFSSLMEVLRPYGAFWLASAFCIFSVLFTLFCVPETKGKTLEQITAHFEGR 477
              |||||||||||:||:||||||||| :|||  ||||||  ||||||:||||:|||||||
Db        420 MAFLVTKEFSSVMEMLRPYGAFWLTAAFCALSVLFTLTVVPETKGRTLEQVTAHFEGR 477

RESULT 4
GTR8_BOVIN
ID   GTR8_BOVIN     STANDARD;      PRT;   334 AA.
AC   P58354;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Solute carrier family 2, facilitated glucose transporter, member 8
DE   (Glucose transporter type 8) (Glucose transporter type X1) (Fragment).
GN   SLC2A8 OR GLUT8 OR GLUTX1.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21482567; PubMed=11599048;
RA   Augustin R., Pocar P., Navarrete-Santos A., Wrenzycki C., Gandolfi F.,
RA   Niemann H., Fischer B.;
RT   "Glucose transporter expression is developmentally regulated in in
RT   vitro derived bovine preimplantation embryos.";
RL   Mol. Reprod. Dev. 60:370-376(2001).
CC   -!- FUNCTION: Insulin-regulated facilitative glucose transporter.
CC       Binds cytochalasin B in a glucose-inhibitable manner. Seems to be
CC       a dual-specific sugar transporter as it is inhibitable by
CC       fructose (By similarity).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Principally
CC       intracellular. May move between intracellular vesicles and the
CC       plasma membrane. The dileucine internalization motif is critical
CC       for intracellular sequestration (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY. GLUCOSE
CC       TRANSPORTERS SUBFAMILY.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; AF321324; AAK69606.1; -.
DR   InterPro; IPR007114; MFS.
```

```
DR   InterPro; IPR005828; Sub_transporter.
DR   InterPro; IPR005829; Sug_transporter.
DR   Pfam; PF00083; sugar_tr; 1.
DR   PROSITE; PS50850; MFS; 1.
DR   PROSITE; PS00216; SUGAR_TRANSPORT_1; 1.
DR   PROSITE; PS00217; SUGAR_TRANSPORT_2; PARTIAL.
KW   Transport; Sugar transport; Transmembrane; Multigene family.
FT   NON_TER       1      1
FT   DOMAIN        1     14       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     15     35       5 (POTENTIAL).
FT   DOMAIN       36     38       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     39     59       6 (POTENTIAL).
FT   DOMAIN       60    113       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    114    134       7 (POTENTIAL).
FT   DOMAIN      135    149       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    150    170       8 (POTENTIAL).
FT   DOMAIN      171    176       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    177    197       9 (POTENTIAL).
FT   DOMAIN      198    224       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    225    245       10 (POTENTIAL).
FT   DOMAIN      246    261       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    262    282       11 (POTENTIAL).
FT   DOMAIN      283    295       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    296    316       12 (POTENTIAL).
FT   DOMAIN      317    334       CYTOPLASMIC (POTENTIAL).
SQ   SEQUENCE   334 AA;  36699 MW;  0EE9B670ADAB71DD CRC64;

  Query Match             62.6%;  Score 1537.5;  DB 1;  Length 334;
  Best Local Similarity   87.7%;  Pred. No. 5.4e-101;
  Matches  293;  Conservative   17;  Mismatches   23;  Indels    1;  Gaps    1;

Qy        145 ISEIAYPAVRGLLGSCVQLMVVVGILLAYLAGWVLEWRWLAVLGCVPPSLMLLLMCFMPE 204
              ||||||| |||||||||||||| ||||||||||||||||||||||| || ||||||||||
Db          1 ISEIAYPEVRGLLGSCVQLMVVTGILLAYLAGWVLEWRWLAVLGCVAPSFMLLLMCFMPE 60

Qy        205 TPRFLLTQHRRQEAMAALRFLWGSEQGWEDPPIGAE-QSFHLALLRQPGIYKPFIIGVSL 263
              ||||||:||: ||||||::||||  ||||:||:||: | ||:| ||:||:|||||||:||
Db         61 TPRFLLSQHKHQEAMAAMQFLWGYAQGWEEPPLGAQHQDFHVAQLRRPGVYKPFIIGISL 120

Qy        264 MAFQQLSGVNAVMFYAETIFEEAKFKDSSLASVVVGVIQVLFTAVAALIMDRAGRRLLLV 323
              |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db        121 MAFQQLSGVNAVMFYAETIFEEAKFKDSSLASVVVGVIQVLFTATAALIMDRAGRRLLLT 180

Qy        324 LSGVVMVFSTSAFGAYFKLTQGGPGNSSHVAISAPVSAQPVDASVGLAWLAVGSMCLFIA 383
              |||||||||||||| |||||:||| ||||| : | || :  | :|||||||||:||||||
Db        181 LSGVVMVFSTSAFGTYFKLTEGGPSNSSHVDLPALVSMEAADTNVGLAWLAVGNMCLFIA 240

Qy        384 GFAVGWGPIPWLLMSEIFPLHVKGVATGICVLTNWLMAFLVTKEFSSLMEVLRPYGAFWL 443
              ||||||||||||||||||||||||||||:|||||| ||||||||||||||||||||||||
Db        241 GFAVGWGPIPWLLMSEIFPLHVKGVATGVCVLTNWFMAFLVTKEFSSLMEVLRPYGAFWL 300

Qy        444 ASAFCIFSVLFTLFCVPETKGKTLEQITAHFEGR 477
              ||||||| ||||| ||||||||||||||||||||
Db        301 ASAFCIFGVLFTLACVPETKGKTLEQITAHFEGR 334


RESULT 5
GTR6_HUMAN
ID   GTR6_HUMAN     STANDARD;      PRT;   507 AA.
AC   Q9UGQ3;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Solute carrier family 2, facilitated glucose transporter, member 6
DE   (Glucose transporter type 6) (Glucose transporter type 9).
GN   SLC2A6 OR GLUT9.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Leukocyte;
```

```
RX   MEDLINE=20427701; PubMed=10970791;
RA   Doege H., Bocianski A., Joost H.-G., Schuermann A.;
RT   "Activity and genomic organization of human glucose transporter 9
RT   (GLUT9), a novel member of the family of sugar-transport facilitators
RT   predominantly expressed in brain and leucocytes.";
RL   Biochem. J. 350:771-776(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Lymphoid;
RA   Young J.M., Woodward K.J., Aziz S., Burley M., Kwiatkowski D.J.,
RA   Povey S.;
RT   "Cloning of a sugar transporter gene, a G-beta subunit like gene and
RT   three novel genes in human chromosome 9q34.";
RL   Submitted (SEP-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Kidney;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC   -!- FUNCTION: Facilitative glucose transporter; binds cytochalasin B
CC       with low affinity.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. The dileucine
CC       internalization motif is critical for intracellular sequestration
CC       (By similarity).
CC   -!- TISSUE SPECIFICITY: Highly expressed in brain, spleen and
CC       peripheral blood leukocytes.
CC   -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY. GLUCOSE
CC       TRANSPORTERS SUBFAMILY.
CC   -!- CAUTION: Has been described as GLUT9 in litterature, but this gene
CC       name has already been used for SLC2A9.
CC
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Y17803; CAB96996.1; -.
DR   EMBL; AJ011372; CAB66155.1; -.
DR   EMBL; BC013740; AAH13740.1; -.
DR   Genew; HGNC:11011; SLC2A6.
DR   MIM; 606813; -.
DR   InterPro; IPR007114; MFS.
DR   InterPro; IPR005828; Sub_transporter.
DR   InterPro; IPR005829; Sug_transporter.
DR   InterPro; IPR003663; Sugar_transpt.
DR   Pfam; PF00083; sugar_tr; 1.
DR   PRINTS; PR00171; SUGRTRNSPORT.
DR   TIGRFAMs; TIGR00879; SP; 1.
DR   PROSITE; PS50850; MFS; 1.
DR   PROSITE; PS00216; SUGAR_TRANSPORT_1; 1.
DR   PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.
```

```
KW   Transport; Sugar transport; Transmembrane; Glycoprotein;
KW   Multigene family.
FT   DOMAIN        1     37       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     38     58       1 (POTENTIAL).
FT   DOMAIN       59     81       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     82    102       2 (POTENTIAL).
FT   DOMAIN      103    111       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    112    132       3 (POTENTIAL).
FT   DOMAIN      133    140       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    141    161       4 (POTENTIAL).
FT   DOMAIN      162    168       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    169    189       5 (POTENTIAL).
FT   DOMAIN      190    194       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    195    215       6 (POTENTIAL).
FT   DOMAIN      216    289       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    290    310       7 (POTENTIAL).
FT   DOMAIN      311    314       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    315    335       8 (POTENTIAL).
FT   DOMAIN      336    339       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    340    360       9 (POTENTIAL).
FT   DOMAIN      361    395       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    396    416       10 (POTENTIAL).
FT   DOMAIN      417    435       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    436    456       11 (POTENTIAL).
FT   DOMAIN      457    462       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    463    483       12 (POTENTIAL).
FT   DOMAIN      484    507       CYTOPLASMIC (POTENTIAL).
FT   SITE          5      6       DILEUCINE INTERNALIZATION MOTIF (BY
FT                                SIMILARITY).
FT   CARBOHYD    370    370       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT    500    500       M -> T (IN REF. 2).
SQ   SEQUENCE   507 AA;  54569 MW;  1FC6E2EB11588460 CRC64;

  Query Match             38.6%;  Score 948;  DB 1;  Length 507;
  Best Local Similarity   42.7%;  Pred. No. 1.9e-59;
  Matches  216;  Conservative   80;  Mismatches  166;  Indels   44;  Gaps   11;

Qy         10 QPLLG------------PPGGSAPRGR-------RVFLAAFAAALGPLSFGFALGYSSPA 50
              :||||            ||     | |       ||||| ||| ||  |||:|| |:||
Db          3 EPLLGAEGPDYDTFPEKPPPSPGDRARVGTLQNKRVFLATFAAVLGNFSFGYALVYTSPV 62

Qy         51 IPSLQRAAPPAPRLDDAAASWFGAVVTLGAAAGGVLGGWLVDRAGRKLSLLLCSVPFVAG 110
              ||:|:|:  |   |  : |||||:| ||||||||:    | |  |||||::  :||  ||
Db         63 IPALERSLDPDLHLTKSQASWFGSVFTLGAAAGGLSAMILNDLLGRKLSIMFSAVPSAAG 122

Qy        111 FAVITAAQDVWMLLGGRLLTGLACGVASLVAPVYISEIAYPAVRGLLGSCVQLMVVVGIL 170
              :|::  |  :|||| || ||| | |: :   |||:|||| | ||| ||:  ||| | | |
Db        123 YALMAGAHGLWMLLLGRTLTGFAGGLTAACIPVYVSEIAPPGVRGALGATPQLMAVFGSL 182

Qy        171 LAYLAGWVLEWRWLAVLGCVPPSLMLLLMCFMPETPRFLLTQHRRQEAMAALRFLWGSEQ 230
                |  | :| |||||| |  |  :|:||: ||| :|||||:: | :||: || :| |::
Db        183 SLYALGLLLPWRWLAVAGEAPVLIMILLLSFMPNSPRFLLSRGRDEEALRALAWLRGTDV 242

Qy        231 G--WEDPPIG---AEQSFHL--ALLRQPGIYKPFIIGVSLMAFQQLSGVNAVMFYAETIF 283
                 ||   |      ||  :  |  | | : :|  : : :   |||:|:  :: | ::||
Db        243 DVHWEFEQIQDNVRRQSSRVSWAEARAPHVCRPITVALLMRLLQQLTGITPILVYLQSIF 302

Qy        284 EEAKF----KDSSLASVVVGVIQVLFTAVAALIMDRAGRRLLLVLSGVVMVFSTSAFGAY 339
              :        ||    : :|| :::|   :||| || |||::|| :|  :|  :    | |
Db        303 DSTAVLLPPKDD---AAIVGAVRLLSVLIAALTMDLAGRKVLLFVSAAIMFAANLTLGLY 359

Qy        340 FKLTQGGPGNSSHVAISAPVS------AQPVDASVG-LAWLAVGSMCLFIAGFAVGWGPI 392
                    ||   |  : :   |      |||: |  | |  : : :  ||| |:|||||||
Db        360 IHF---GPRPLSPNSTAGLESESWGDLAQPLAAPAGYLTLVPLLATMLFIMGYAVGWGPI 416

Qy        393 PWLLMSEIFPLHVKGVATGICVLTNWLMAFLVTKEFSSLMEVLRPYGAFWLASAFCIFSV 452
               ||||||: ||  :|||:|:||| :|| ||::|| |  ::        |:  :| |: |:
Db        417 TWLLMSEVLPLRARGVASGLCVLASWLTAFVLTKSFLPVVSTFGLQVPFFFFAAICLVSL 476

Qy        453 LFTLFCVPETKGKTLEQITAHFE-GR 477
              :||  |||||||::|||| : |  ||
Db        477 VFTGCCVPETKGRSLEQIESFFRMGR 502
```

```
RESULT 6
CSBC_BACSU
ID   CSBC_BACSU     STANDARD;      PRT;   461 AA.
AC   P46333; O32289;
DT   01-NOV-1995 (Rel. 32, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Probable metabolite transport protein csbC.
GN   CSBC OR SS92BR OR BSU39810.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168 / BGSC1A1;
RX   MEDLINE=96093926; PubMed=7584049;
RA   Yoshida K.-I., Seki S., Fujimura M., Miwa Y., Fujita Y.;
RT   "Cloning and sequencing of a 36-kb region of the Bacillus subtilis
RT   genome between the gnt and iol operons.";
RL   DNA Res. 2:61-69(1995).
RN   [2]
RP   REVISIONS.
RA   Fujita Y., Shibayama T., Ishio I., Aoyama D., Yoshida K.-I.;
RL   Submitted (JUN-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=98044033; PubMed=9384377;
RA   Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA   Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA   Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA   Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA   Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA   Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA   Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA   Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA   Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA   Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA   Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA   Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA   Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA   Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA   Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA   Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA   Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA   Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA   Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA   Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA   Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA   Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA   Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA   Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA   Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA   Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA   Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA   Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT   "The complete genome sequence of the Gram-positive bacterium Bacillus
RT   subtilis.";
RL   Nature 390:249-256(1997).
RN   [4]
RP   CHARACTERIZATION.
RX   MEDLINE=99303315; PubMed=10376822;
RA   Akbar S., Lee S.Y., Boylan S.A., Price C.W.;
RT   "Two genes from Bacillus subtilis under the sole control of the
RT   general stress transcription factor sigmaB.";
RL   Microbiology 145:1069-1078(1999).
CC   -!- FUNCTION: Could serve either a nutritional or an osmotic
CC       protection function.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC   -!- SIMILARITY: Belongs to the sugar transporter family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; AB005554; BAA21604.1; -.
DR   EMBL; Z99124; CAB16017.1; -.
DR   PIR; D70073; D70073.
DR   SubtiList; BG11360; csbC.
DR   InterPro; IPR007114; MFS.
DR   InterPro; IPR005828; Sub_transporter.
DR   InterPro; IPR005829; Sug_transporter.
DR   InterPro; IPR003663; Sugar_transpt.
DR   Pfam; PF00083; sugar_tr; 1.
DR   PRINTS; PR00171; SUGRTRNSPORT.
DR   TIGRFAMs; TIGR00879; SP; 1.
DR   PROSITE; PS50850; MFS; 1.
DR   PROSITE; PS00216; SUGAR_TRANSPORT_1; 2.
DR   PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.
KW   Transport; Transmembrane; Complete proteome.
FT   DOMAIN        1     14       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     15     35       1 (POTENTIAL).
FT   DOMAIN       36     38       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     39     59       2 (POTENTIAL).
FT   DOMAIN       60     76       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     77     97       3 (POTENTIAL).
FT   DOMAIN       98    104       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    105    125       4 (POTENTIAL).
FT   DOMAIN      126    139       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    140    160       5 (POTENTIAL).
FT   DOMAIN      161    163       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    164    184       6 (POTENTIAL).
FT   DOMAIN      185    241       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    242    262       7 (POTENTIAL).
FT   DOMAIN      263    280       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    281    301       8 (POTENTIAL).
FT   DOMAIN      302    308       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    309    329       9 (POTENTIAL).
FT   DOMAIN      330    341       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    342    362       10 (POTENTIAL).
FT   DOMAIN      363    378       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    379    399       11 (POTENTIAL).
FT   DOMAIN      400    402       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    403    423       12 (POTENTIAL).
FT   DOMAIN      424    461       CYTOPLASMIC (POTENTIAL).
FT   CONFLICT    400    401       SA -> RP (IN REF. 1).
SQ   SEQUENCE   461 AA;  50140 MW;  AD59CB5DD6372FE0 CRC64;

  Query Match             24.7%;  Score 607;  DB 1;  Length 461;
  Best Local Similarity   31.1%;  Pred. No. 1.5e-35;
  Matches  146;  Conservative   85;  Mismatches  192;  Indels   46;  Gaps    7;

Qy          25 RVFLAAFAAALGPLSFGFALGYSSPAIPSLQRAAPPAPRLDDAAASWFGAVVTLGAAAGG 84
               | ::  |  ||| | :|:  |  | |:  :    |     :    |    :: |||  |
Db           6 RKYMIYFFGALGGLLYGYDTGVISGALLFINNDIPLTTLTEGLVVS----MLLLGAIFGS 61

Qy          85 VLGGWLVDRAGRKLSLLLCSVPFVAGFAVITAAQDVWMLLGGRLLTGLACGVASLVAPVY 144
                | |   || ||:  : : |: |: |      :| : ||:  |:: ||| | :: : |||
Db          62 ALSGTCSDRWGRRKVVFVLSIIFIIGALACAFSQTIGMLIASRVILGLAVGGSTALVPVY 121

Qy         145 ISEIAYPAVRGLLGSCVQLMVVVGILLAYLAGWVL----EWRWLAVLGCVPPSLMLLLMC 200
               :||:|   :|| ||:   ||:| ||||||:  ::      |||:  |  ||  |:|: :
Db         122 LSEMAPTKIRGTLGTMNNLMIVTGILLAYIVNYLFTPFEAWRWMVGLAAVPAVLLLIGIA 181

Qy         201 FMPETPRFLLTQHRRQEAMAALRF----------LWGSEQGWEDPPIGAEQSFHLALLRQ 250
               ||||:||:|: :   :||   :            |   :||  :     ::   | :|:
Db         182 FMPESPRWLVKRGSEEEARRIMNITHDPKDIEMELAEMKQGEAE-----KKETTLGVLKA 236

Qy         251 PGIYKPFIIGVSLMAFQQLSGVNAVMFYAETIFEEAKF--KDSSLASVVVGVIQVLFTAV 308
                 |    :||| |  |||  |:| |::|| ||| :|      |:| :: :|:: |:
Db         237 KWIRPMLLIGVGLAIFQQAVGINTVIYYAPTIFTKAGLGTSASALGTMGIGILNVIMCIT 296

Qy         309 AALIMDRAGRRLLLVLSGVVMVFSTSAFGAYFKLTQGGPGNSSHVAISAPVSAQPVDASV 368
               | :::|| ||: ||:   | :  | :|      || |                     |
Db         297 AMILIDRVGRKKLLIWGSVGITLSLAALSGVL-LTLG--------------------LSA 335

Qy         369 GLAWLAVGSMCLFIAGFAVGWGPIPWLLMSEIFPLHVKGVATGICVLTNWLMAFLVTKEF 428
                 ||: |  : ::|  :   |||: |:|| |:||   :| |||   |       :|:  |
Db         336 STAWMTVVFLGVYIVFYQATWGPVVWVLMPELFPSKARGAATGFTTLVLSAANLIVSLVF 395

Qy         429 SSLMEVLRPYGAFWLASAFCIFSVLFTLFCVPETKGKTLEQITAHFEGR 477
                 ::  :     | : |  |: |  |  : |||||||:||:| |  : |
Db         396 PLMLSAMGIAWVFMVFSVICLLSFFFAFYMVPETKGKSLEEIEASLKKR 444


RESULT 7
ARAE_KLEOX
ID   ARAE_KLEOX     STANDARD;      PRT;   472 AA.
AC   P45598;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Arabinose-proton symporter (Arabinose transporter).
GN   ARAE.
OS   Klebsiella oxytoca.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Klebsiella.
OX   NCBI_TaxID=571;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=8017;
RX   MEDLINE=95394866; PubMed=7665532;
RA   Shatwell K.P., Charalambous B.M., McDonald T.P., Henderson P.J.F.;
RT   "Cloning, sequencing, and expression of the araE gene of Klebsiella
RT   oxytoca 8017, which encodes arabinose-H+ symport activity.";
RL   J. Bacteriol. 177:5379-5380(1995).
CC   -!- FUNCTION: Uptake of arabinose across the boundary membrane with
CC       the concomitant export of a proton (symport system).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.
CC   -!- SIMILARITY: Belongs to the sugar transporter family.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; X79598; CAA56110.1; -.
DR   PIR; S47089; S47089.
DR   InterPro; IPR007114; MFS.
DR   InterPro; IPR005828; Sub_transporter.
DR   InterPro; IPR005829; Sug_transporter.
DR   InterPro; IPR003663; Sugar_transpt.
DR   Pfam; PF00083; sugar_tr; 1.
DR   PRINTS; PR00171; SUGRTRNSPORT.
DR   TIGRFAMs; TIGR00879; SP; 1.
DR   PROSITE; PS50850; MFS; 1.
DR   PROSITE; PS00216; SUGAR_TRANSPORT_1; 1.
DR   PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.
KW   Transport; Sugar transport; Transmembrane; Inner membrane; Symport.
FT   DOMAIN        1     29       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     30     50       1 (POTENTIAL).
FT   DOMAIN       51     63       PERIPLASMIC (POTENTIAL).
FT   TRANSMEM     64     84       2 (POTENTIAL).
FT   DOMAIN       85     91       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     92    112       3 (POTENTIAL).
FT   DOMAIN      113    114       PERIPLASMIC (POTENTIAL).
FT   TRANSMEM    115    135       4 (POTENTIAL).
FT   DOMAIN      136    154       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    155    175       5 (POTENTIAL).
FT   DOMAIN      176    178       PERIPLASMIC (POTENTIAL).
FT   TRANSMEM    179    199       6 (POTENTIAL).
```

```
FT   DOMAIN       200    257       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     258    278       7 (POTENTIAL).
FT   DOMAIN       279    297       PERIPLASMIC (POTENTIAL).
FT   TRANSMEM     298    318       8 (POTENTIAL).
FT   DOMAIN       319    325       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     326    346       9 (POTENTIAL).
FT   DOMAIN       347    361       PERIPLASMIC (POTENTIAL).
FT   TRANSMEM     362    382       10 (POTENTIAL).
FT   DOMAIN       383    404       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     405    425       11 (POTENTIAL).
FT   DOMAIN       426    427       PERIPLASMIC (POTENTIAL).
FT   TRANSMEM     428    446       12 (POTENTIAL).
FT   DOMAIN       447    472       CYTOPLASMIC (POTENTIAL).
SQ   SEQUENCE   472 AA;  51732 MW;  410021E1BEE3D96E CRC64;

  Query Match             24.7%;  Score 606.5;  DB 1;  Length 472;
  Best Local Similarity   33.7%;  Pred. No. 1.7e-35;
  Matches  153;  Conservative   84;  Mismatches  182;  Indels   35;  Gaps    8;

Qy          30 AFAAALGPLSFGFALGYSSPAIPSLQRAAPPAPRLDDAAASWFGAVVTLGAAAGGVLGGW 89
               : |||:  | ||  :|  : |:| :      : || :    |  : : |||| | :  ||
Db          26 SIAAAVAGLLFGLDIGVIAGALPFITDHFVLSSRLQE----WVVSSMMLGAAIGALFNGW 81

Qy          90 LVDRAGRKLSLLLCSVPFVAGFAVITAAQDVWMLLGGRLLTGLACGVASLVAPVYISEIA 149
               |  | ||| ||:: :| |||      |  | |||  |:: |:| |:|| ||:|:||:|
Db          82 LSFRLGRKYSLMVGAVLFVAGSVGSAFATSVEMLLVARIVLGVAVGIASYTAPLYLSEMA 141

Qy         150 YPAVRGLLGSCVQLMVVVGILLAYLAGWVLEW--RWLAVLG--CVPPSLMLLLMCFMPET 205
                  ||| : |  |||| :||::|:|:     :   | |:||   :|  ::::|: |:| :
Db         142 SENVRGKMISMYQLMVTLGIVMAFLSDTAFSYSGNWRAMLGVLALPAVVLIILVIFLPNS 201

Qy         206 PRFLLTQHRRQEAMAALRFLWGSEQGWEDPPIGAEQSFHL-----ALLR-QPGIYKPFII 259
               ||:|  : |  ||   || |  : :   |      :|  |     || :    : :   :
Db         202 PRWLAEKGRHVEAEEVLRMLRDTSEKARDELNEIRESLKLKQGGWALFKVNRNVRRAVFL 261

Qy         260 GVSLMAFQQLSGVNAVMFYAETIFEEAKF---KDSSLASVVVGVIQVLFTAVAALIMDRA 316
               |: | | || :|:| :|:||  ||: | |   :   :|::|||:  :  |   :|   :|:|
Db         262 GMLLQAMQQFTGMNIIMYYAPRIFKMAGFTTTEQQMVATLVVGLTFMFATFIAVFTVDKA 321

Qy         317 GRRLLLVLSGVVMVFSTSAFGAYFKLTQGGPGNSSHVAISAPVSAQPVDASVGLAWLAVG 376
               ||:  | :   ||   |   |  : | |    |                  || ||:||:||
Db         322 GRKPALKIGFSVMAIGTLVLG--YCLMQFDNGT----------------ASSGLSWLSVG 363

Qy         377 SMCLFIAGFAVGWGPIPWLLMSEIFPLHVKGVATGICVLTNWLMAFLVTKEFSSLMEVLR 436
                  : |||:|:   |: |:| ||| ||  :         |||:   ::   | :|:: :
Db         364 MTMMCIAGYAMSAAPVVWILCSEIQPLKCRDFGITCSTTTNWVSNMIIGATFLTLLDAIG 423

Qy         437 PYGAFWLASAFCIFSVLFTLFCVPETKGKTLEQI 470
                 | ||| :|  :  :  | : :||||  ||| |
Db         424 AAGTFWLYTALNVAFIGVTFWLIPETKNVTLEHI 457


RESULT 8
ARAE_ECOLI
ID   ARAE_ECOLI     STANDARD;      PRT;   472 AA.
AC   P09830; Q46937;
DT   01-MAR-1989 (Rel. 10, Created)
DT   01-MAR-1989 (Rel. 10, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Arabinose-proton symporter (Arabinose transporter).
GN   ARAE OR B2841 OR Z4161 OR ECS3698.
OS   Escherichia coli, and
OS   Escherichia coli O157:H7.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=562, 83334;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87115869; PubMed=3543693;
RA   Maiden M.C.J., Davis E.O., Baldwin S.A., Moore D.C.M.,
RA   Henderson P.J.F.;
RT   "Mammalian and bacterial sugar transport proteins are homologous.";
RL   Nature 325:641-643(1987).
```

```
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / JM2433;
RX   MEDLINE=88228015; PubMed=2836407;
RA   Maiden M.C.J., Jones-Mortimer M.C., Henderson P.J.F.;
RT   "The cloning, DNA sequence, and overexpression of the gene araE
RT   coding for arabinose-proton symport in Escherichia coli K12.";
RL   J. Biol. Chem. 263:8003-8010(1988).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX   MEDLINE=21074935; PubMed=11206551;
RA   Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA   Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA   Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA   Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA   Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA   Welch R.A., Blattner F.R.;
RT   "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL   Nature 409:529-533(2001).
RN   [5]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / RIMD 0509952;
RX   MEDLINE=21156231; PubMed=11258796;
RA   Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA   Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA   Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA   Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT   "Complete genome sequence of enterohemorrhagic Escherichia coli
RT   O157:H7 and genomic comparison with a laboratory strain K-12.";
RL   DNA Res. 8:11-22(2001).
RN   [6]
RP   PRELIMINARY SEQUENCE OF 1-28 FROM N.A.
RX   MEDLINE=84114868; PubMed=6319708;
RA   Stoner C., Schleif R.F.;
RT   "The araE low affinity L-arabinose transport promoter. Cloning,
RT   sequence, transcription start site and DNA binding sites of
RT   regulatory proteins.";
RL   J. Mol. Biol. 171:369-381(1983).
CC   -!- FUNCTION: Uptake of arabinose across the boundary membrane with
CC       the concomitant export of a proton (symport system).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.
CC   -!- SIMILARITY: Belongs to the sugar transporter family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; J03732; AAA23469.1; -.
DR   EMBL; X00272; CAA25075.1; -.
DR   EMBL; U29581; AAB40488.1; -.
DR   EMBL; AE000368; AAC75880.1; -.
DR   EMBL; AE005513; AAG57953.1; -.
DR   EMBL; AP002563; BAB37121.1; -.
DR   PIR; B26430; B26430.
DR   PIR; B91091; B91091.
DR   PIR; E85936; E85936.
DR   EcoGene; EG10056; araE.
DR   InterPro; IPR007114; MFS.
```

```
DR   InterPro; IPR005828; Sub_transporter.
DR   InterPro; IPR005829; Sug_transporter.
DR   InterPro; IPR003663; Sugar_transpt.
DR   Pfam; PF00083; sugar_tr; 1.
DR   PRINTS; PR00171; SUGRTRNSPORT.
DR   TIGRFAMs; TIGR00879; SP; 1.
DR   PROSITE; PS50850; MFS; 1.
DR   PROSITE; PS00216; SUGAR_TRANSPORT_1; 1.
DR   PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.
KW   Transport; Sugar transport; Transmembrane; Inner membrane; Symport;
KW   Complete proteome.
FT   DOMAIN        1     29       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     30     50       1 (POTENTIAL).
FT   DOMAIN       51     63       PERIPLASMIC (POTENTIAL).
FT   TRANSMEM     64     84       2 (POTENTIAL).
FT   DOMAIN       85     91       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     92    112       3 (POTENTIAL).
FT   DOMAIN      113    114       PERIPLASMIC (POTENTIAL).
FT   TRANSMEM    115    135       4 (POTENTIAL).
FT   DOMAIN      136    154       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    155    175       5 (POTENTIAL).
FT   DOMAIN      176    178       PERIPLASMIC (POTENTIAL).
FT   TRANSMEM    179    199       6 (POTENTIAL).
FT   DOMAIN      200    257       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    258    278       7 (POTENTIAL).
FT   DOMAIN      279    297       PERIPLASMIC (POTENTIAL).
FT   TRANSMEM    298    318       8 (POTENTIAL).
FT   DOMAIN      319    325       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    326    346       9 (POTENTIAL).
FT   DOMAIN      347    361       PERIPLASMIC (POTENTIAL).
FT   TRANSMEM    362    382       10 (POTENTIAL).
FT   DOMAIN      383    404       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    405    425       11 (POTENTIAL).
FT   DOMAIN      426    427       PERIPLASMIC (POTENTIAL).
FT   TRANSMEM    428    446       12 (POTENTIAL).
FT   DOMAIN      447    472       CYTOPLASMIC (POTENTIAL).
FT   CONFLICT     26     28       SVA -> YDR (IN REF. 6).
SQ   SEQUENCE   472 AA;  51684 MW;  411990A441D44393 CRC64;

  Query Match             24.5%;  Score 603;  DB 1;  Length 472;
  Best Local Similarity   33.3%;  Pred. No. 2.9e-35;
  Matches  158;  Conservative   84;  Mismatches  197;  Indels   36;  Gaps    9;

Qy            9 TQPLLGPPGGSAPRGRRVFLAAFAAALGPLSFGFALGYSSPAIPSLQRAAPPAPRLDDAA 68
                |:  | |      |   :|::  |||:  | ||  :|  : |:| :        || :
Db            6 TESALTPRSLRDTRRMNMFVSV-AAAVAGLLFGLDIGVIAGALPFITDHFVLTSRLQE-- 62

Qy           69 ASWFGAVVTLGAAAGGVLGGWLVDRAGRKLSLLLCSVPFVAGFAVITAAQDVWMLLGGRL 128
                  |  : : |||| | :  |||  | |||  ||:  :: || |      |  | ||:  |:
Db           63 --WVVSSMMLGAAIGALFNGWLSFRLGRKYSLMAGAILFVLGSIGSAFATSVEMLIAARV 120

Qy          129 LTGLACGVASLVAPVYISEIAYPAVRGLLGSCVQLMVVVGILLAYLAGWVLEW--RWLAV 186
                : |:| |:||  ||:|:||:|   ||| : |  |||| :||:||:|:     :   | |:
Db          121 VLGIAVGIASYTAPLYLSEMASENVRGKMISMYQLMVTLGIVLAFLSDTAFSYSGNWRAM 180

Qy          187 LG--CVPPSLMLLLMCFMPETPRFLLTQHRRQEAMAALRFLWGSEQGWEDPPIGAEQSFH 244
                ||   :|  |:::|: |:| :||:|  : |  ||   || |  : :   :      :|
Db          181 LGVLALPAVLLIILVVFLPNSPRWLAEKGRHIEAEEVLRMLRDTSEKAREELNEIRESLK 240

Qy          245 L-----ALLR-QPGIYKPFIIGVSLMAFQQLSGVNAVMFYAETIFEEAKF---KDSSLAS 295
                |     || :    : :   :|: | | || :|:| :|:||  ||: | |   :   :|:
Db          241 LKQGGWALFKINRNVRRAVFLGMLLQAMQQFTGMNIIMYYAPRIFKMAGFTTTEQQMIAT 300

Qy          296 VVVGVIQVLFTAVAALIMDRAGRRLLLVLSGVVMVFSTSAFGAYFKLTQGGPGNSSHVAI 355
                :|||:  :  | :|   :|:|||:  | :   ||   |   |  : | |   |
Db          301 LVVGLTFMFATFIAVFTVDKAGRKPALKIGFSVMALGTLVLG--YCLMQFDNGT------ 352

Qy          356 SAPVSAQPVDASVGLAWLAVGSMCLFIAGFAVGWGPIPWLLMSEIFPLHVKGVATGICVL 415
                          || ||:||:||   : |||:|:   |: |:| ||| ||  :
Db          353 ----------ASSGLSWLSVGMTMMCIAGYAMSAAPVVWILCSEIQPLKCRDFGITCSTT 402

Qy          416 TNWLMAFLVTKEFSSLMEVLRPYGAFWLASAFCIFSVLFTLFCVPETKGKTLEQI 470
                |||:   ::   | :|:: :   | ||| :|  |  |  | : :||||  ||| |
Db          403 TNWVSNMIIGATFLTLLDSIGAAGTFWLYTALNIAFVGITFWLIPETKNVTLEHI 457

RESULT 9
GALP_ECOLI
ID   GALP_ECOLI     STANDARD;      PRT;   464 AA.
AC   P37021;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Galactose-proton symporter (Galactose transporter).
GN   GALP OR B2943 OR C3529.
OS   Escherichia coli, and
OS   Escherichia coli O6.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=562, 217992;
RN   [1]
RP   SEQUENCE FROM N.A., AND CHARACTERIZATION.
RA   Roberts P.E.;
RL   Thesis (1992), University of Cambridge, U.K.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX   MEDLINE=22388234; PubMed=12471157;
RA   Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA   Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA   Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA   Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT   "Extensive mosaic structure revealed by the complete genome sequence
RT   of uropathogenic Escherichia coli.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
CC   -!- FUNCTION: UPTAKE OF GALACTOSE ACROSS THE BOUNDARY MEMBRANE WITH
CC       THE CONCOMITANT EXPORT OF A PROTON (SYMPORT SYSTEM).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.
CC   -!- SIMILARITY: Belongs to the sugar transporter family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U28377; AAA69110.1; -.
DR   EMBL; AE000377; AAC75980.1; -.
DR   EMBL; AE016766; AAN81977.1; ALT_INIT.
DR   PIR; F65079; F65079.
DR   EcoGene; EG12148; galP.
DR   InterPro; IPR007114; MFS.
DR   InterPro; IPR005828; Sub_transporter.
DR   InterPro; IPR005829; Sug_transporter.
DR   InterPro; IPR003663; Sugar_transpt.
DR   Pfam; PF00083; sugar_tr; 1.
DR   PRINTS; PR00171; SUGRTRNSPORT.
DR   TIGRFAMs; TIGR00879; SP; 1.
DR   PROSITE; PS50850; MFS; 1.
DR   PROSITE; PS00216; SUGAR_TRANSPORT_1; 1.
DR   PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.
KW   Transport; Sugar transport; Transmembrane; Inner membrane; Symport;
KW   Complete proteome.
FT   DOMAIN        1     15       CYTOPLASMIC (POTENTIAL).
```

```
FT   TRANSMEM     16     36       1 (POTENTIAL).
FT   DOMAIN       37     56       PERIPLASMIC (POTENTIAL).
FT   TRANSMEM     57     77       2 (POTENTIAL).
FT   DOMAIN       78     84       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     85    105       3 (POTENTIAL).
FT   DOMAIN      106    112       PERIPLASMIC (POTENTIAL).
FT   TRANSMEM    113    133       4 (POTENTIAL).
FT   DOMAIN      134    139       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    140    160       5 (POTENTIAL).
FT   DOMAIN      161    171       PERIPLASMIC (POTENTIAL).
FT   TRANSMEM    172    192       6 (POTENTIAL).
FT   DOMAIN      193    250       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    251    271       7 (POTENTIAL).
FT   DOMAIN      272    290       PERIPLASMIC (POTENTIAL).
FT   TRANSMEM    291    311       8 (POTENTIAL).
FT   DOMAIN      312    321       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    322    342       9 (POTENTIAL).
FT   DOMAIN      343    351       PERIPLASMIC (POTENTIAL).
FT   TRANSMEM    352    372       10 (POTENTIAL).
FT   DOMAIN      373    394       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    395    415       11 (POTENTIAL).
FT   DOMAIN      416    416       PERIPLASMIC (POTENTIAL).
FT   TRANSMEM    417    437       12 (POTENTIAL).
FT   DOMAIN      438    464       CYTOPLASMIC (POTENTIAL).
SQ   SEQUENCE   464 AA;  50982 MW;  07E08935BD8E3F8E CRC64;

  Query Match             24.1%;  Score 591.5;  DB 1;  Length 464;
  Best Local Similarity   32.4%;  Pred. No. 1.8e-34;
  Matches  155;  Conservative   83;  Mismatches  175;  Indels   65;  Gaps   11;

Qy         27 FLAAFAALGPLSFGFALGYSSPAIPSLQRAAPPAPRLDDAAASWFGAVVTLGAAAGGVL 86
              |   | |||  | ||  :|  : |:| :        ::      |  : :  ||| | |
Db         16 FFVCFLAALAGLLFGLDIGVIAGALPFIADEF----QITSHTQEWVVSSMMFGAAVGAVG 71

Qy         87 GGWLVDRAGRKLSLLLCSVPFVAGFAVITAAQDVWMLLGGRLLTGLACGVASLVAPVYIS 146
               |||  : ||| ||:: :: ||||     || :| :|:  |:| ||| ||||  ||:|:|
Db         72 SGWLSFKLGRKKSLMIGAILFVAGSLFSAAAPNVEVLILSRVLLGLAVGVASYTAPLYLS 131

Qy        147 EIAYPAVRGLLGSCVQLMVVVGILLAYLA----GWVLEWRWLAVLGCVPPSLMLLLMCFM 202
              |||   :|| : |  |||: :||| |||:     :   |||:  :  :|  |:|: : |:
Db        132 EIAPEKIRGSMISMYQLMITIGILGAYLSDTAFSYTGAWRWMLGVIIIPAILLLIGVFFL 191

Qy        203 PETPRFLLTQHRRQEAMAAL-------------------RFLWGSEQGWEDPPIGAEQSFH 244
              |::||:   : |  :|   |                    |   : ||
Db        192 PDSPRWFAAKRRFVDAERVLLRLRDTSAEAKRELDEIRESLQVKQSGW------------ 239

Qy        245 LALLRQPGIYKPFI-IGVSLMAFQQLSGVNAVMFYAETIFEEAKFKDSS---LASVVVGV 300
               || ::   ::  : :|| |   || :|:| :|:||  ||| | : :::     :|:||:
Db        240 -ALFKENSNFRRAVFLGVLLQVMQQFTGMNVIMYYAPKIFELAGYTNTTEQMWGTVIVGL 298

Qy        301 IQVLFTAVAALIMDRAGRRLLLVLSGVVMVFSTSAFGAYFKLTQGGPGNSSHVAISAPVS 360
                || | :|  ::|| ||:  | |  :||     | |       |  |   |: | :| |
Db        299 TNVLATFIAIGLVDRWGRKPTLTLGFLVM-----AAG------MGVLGTMMHIGIHSP-S 346

Qy        361 AQPVDASVGLAWLAVGSMCLFIAGFAVGWGPIPWLLMSEIFPLHVKGVATGICVLTNWLM 420
              ||         : |:  : :|| |||:  ||: |:| ||| ||  :         |||:
Db        347 AQ---------YFAIAMLLMFIVGFAMSAGPLIWVLCSEIQPLKGRDFGITCSTATNWIA 397

Qy        421 AFLVTKEFSSLMEVLRPYGAFWLASAFCIFSVLFTLFCVPETKGKTLEQITAH-FEGR 477
                :|   | :::  |     ||: :|  :  :| ||: |||||  :|| |  :  :||
Db        398 NMIVGATFLTMLNTLGNANTFWVYAALNVLFILLTLWLVPETKHVSLEHIERNLMKGR 455


RESULT 10
ITR1_SCHPO
ID   ITR1_SCHPO     STANDARD;      PRT;   575 AA.
AC   Q10286;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Myo-inositol transporter 1.
GN   ITR1 OR SPAC4F8.15 OR SPAC7D4.01.
OS   Schizosaccharomyces pombe (Fission yeast).
```

```
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=968 h90;
RX   MEDLINE=98228265; PubMed=9560432;
RA   Niederberger C., Graub R., Schweingruber A.-M., Fankhauser H.,
RA   Rusu M., Poitelea M., Edenharter L., Schweingruber M.E.;
RT   "Exogenous inositol and genes responsible for inositol transport are
RT   required for mating and sporulation in Shizosaccharomyces pombe.";
RL   Curr. Genet. 33:255-261(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RX   MEDLINE=21848401; PubMed=11859360;
RA   Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA   Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA   Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA   Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA   Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA   Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA   James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA   Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA   Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbinowitsch E.,
RA   Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA   Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA   Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA   Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA   Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA   Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA   Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA   Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA   Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA   Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA   Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA   Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA   Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA   Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT   "The genome sequence of Schizosaccharomyces pombe.";
RL   Nature 415:871-880(2002).
CC   -!- FUNCTION: TRANSPORTER FOR MYO-INOSITOL.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: Belongs to the sugar transporter family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X98622; CAA67211.1; -.
DR   EMBL; Z98530; CAB11061.1; -.
DR   EMBL; Z99532; CAB16718.1; -.
DR   PIR; T43400; T43400.
DR   GeneDB_SPombe; SPAC4F8.15; -.
DR   GO; GO:0000747; P:conjugation with cellular fusion; ISS.
DR   GO; GO:0006629; P:lipid metabolism; ISS.
DR   GO; GO:0007165; P:signal transduction; ISS.
DR   GO; GO:0007151; P:sporulation (sensu Saccharomyces); ISS.
DR   InterPro; IPR007114; MFS.
DR   InterPro; IPR005828; Sub_transporter.
DR   InterPro; IPR005829; Sug_transporter.
DR   InterPro; IPR003663; Sugar_transpt.
DR   Pfam; PF00083; sugar_tr; 1.
DR   PRINTS; PR00171; SUGRTRNSPORT.
DR   TIGRFAMs; TIGR00879; SP; 1.
DR   PROSITE; PS50850; MFS; 1.
DR   PROSITE; PS00216; SUGAR_TRANSPORT_1; 2.
```

```
DR   PROSITE; PS00217; SUGAR_TRANSPORT_2; FALSE_NEG.
KW   Transmembrane; Sugar transport; Glycoprotein.
FT   DOMAIN        1     86       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     87    107       1 (POTENTIAL).
FT   DOMAIN      108    129       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    130    150       2 (POTENTIAL).
FT   DOMAIN      151    156       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    157    177       3 (POTENTIAL).
FT   DOMAIN      178    186       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    187    207       4 (POTENTIAL).
FT   DOMAIN      208    215       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    216    236       5 (POTENTIAL).
FT   DOMAIN      237    246       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    247    267       6 (POTENTIAL).
FT   DOMAIN      268    349       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    350    370       7 (POTENTIAL).
FT   DOMAIN      371    376       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    377    397       8 (POTENTIAL).
FT   DOMAIN      398    400       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    401    421       9 (POTENTIAL).
FT   DOMAIN      422    441       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    442    462       10 (POTENTIAL).
FT   DOMAIN      463    486       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    487    507       11 (POTENTIAL).
FT   DOMAIN      508    510       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    511    531       12 (POTENTIAL).
FT   DOMAIN      532    575       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD    432    432       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   575 AA;  62757 MW;  3B7C5EFF86C596AE CRC64;

  Query Match             23.8%;  Score 584.5;  DB 1;  Length 575;
  Best Local Similarity   29.8%;  Pred. No. 7.1e-34;
  Matches  147;  Conservative   90;  Mismatches  201;  Indels   55;  Gaps   11;

Qy          5 DPEETQPLLGPPGGSAPRGRRVFLAAFAAALGPLSFGFALGYSSPAIPSLQRAAPPAPRL 64
              |  | ||:           : |:: |||| :| | ||:  |  || :  :  :      |
Db         76 DTVEDQPV----------SKWVWVLAFAAGIGGLLFGYDTGVISGALVVIGTSL-GGHEL 124

Qy         65 DDAAASWFGAVVTLGAAAGGVLGGWLVDRAGRKLSLLLCSVPFVAGFAVITAAQDVWMLL 124
               :    :  :  :|||  ||:: | | |  |||  : : |:  : |  |   |  :| ::
Db        125 TNGGKEFITSATSLGALLGGIIAGALADFFGRKPVIAIASIIIIVGSIVQVTAHHLWHMI 184

Qy        125 GGRLLTGLACGVASLVAPVYISEIAYPAVRGLLGSCVQLMVVVGILLAYLAGWVLE---- 180
               || : |   |:|||: |:|:||||   :|| |     |::  | ::||      |
Db        185 VGRFVIGWGVGIASLIIPLYLSEIAPSKIRGRLVIIYVLLITAGQVIAYGIDTAFEHVHN 244

Qy        181 -WRWLAVLGCVPPSLMLLLMCFMPETPRFLLTQHRRQEAMAALRFLWGS----------- 228
               |||:  |  || :  | ::  ::||:|| |: : | |||   |  :: :
Db        245 GWRWMVGLAMVPAAFQLFILIWLPESPRLLVKKERSQEAYNTLARIYPTAHPYEIKTKLY 304

Qy        229 --EQGWEDPPIGAE-----QSFHLALLRQPGIYKPFIIGVSLMAFQQLSGVNAVMFYAET 281
                ::|  ||  |:      ::|   |   |  ::  |:   | | ||||| |::|::: |
Db        305 LIQEGVRDPFSGSRWQKIVKTFK-ELYFNPSNFRALILACGLQAMQQLSGFNSLMYFSST 363

Qy        282 IFEEAKFKDSSLASVVVGVIQVLFTAVAALIMDRAGRRLLLVLS--GVVMVFSTSAFGAY 339
              |||   | : :   :::     :|| ||  ::|  |||:||:|:|  |::      |   :
Db        364 IFEVVGFNNPTATGLIIAATNFVFTIVAFGVIDFFGRRILLLLTVWGMIAALIVCAVAFH 423

Qy        340 FKLTQGGPGNSSHVAISAPVSAQPVDASVGLAWLAVG--SMCLFIAGFAVGWGPIPWLLM 397
              | | :   || :              :    ||  |   || :::| :| | | :||
Db        424 F-LPKDENGNYT---------------SGQSNAWAIVVLISMIVYVASYASGLGNLPW-QQ 467

Qy        398 SEIFPLHVKGVATGICVLTNWLMAFLVTKEFSSLMEVLRPYGAFWLASAFCIFSVLFTLF 457
              ||:||: |:|: ||:    ||     :   | :||  : | | | |    |    |  ||
Db        468 SELFPMSVRGLGTGMSTAVNWAGNLGIGASFLTLMSEITPTGTFALYGGLCFLGWLGALF 527

Qy        458 CVPETKGKTLEQI 470
              | |:    |:|:|
Db        528 CYPDLTDYTIEEI 540


RESULT 11
GTR3_BOVIN
```

```
ID   GTR3_BOVIN     STANDARD;      PRT;   494 AA.
AC   P58352;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Solute carrier family 2, facilitated glucose transporter, member 3
DE   (Glucose transporter type 3, brain).
GN   SLC2A3 OR GLUT3.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Augustin R., Fischer B.;
RT   "Glucose transporter expression during bovine preimplantation embryo
RT   development.";
RL   Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 385-494 FROM N.A.
RC   TISSUE=Brain;
RA   Augustin R., Navarrete-Santos A., Fischer B.;
RT   "Expression of glucose transporters amongst ruminants.";
RL   Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Facilitative glucose transporter. Probably a neuronal
CC       glucose transporter.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: Belongs to the sugar transporter family. Glucose
CC       transporter subfamily.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AY033938; AAK70222.1; -.
DR   EMBL; AF308829; AAK63202.1; -.
DR   InterPro; IPR007114; MFS.
DR   InterPro; IPR005828; Sub_transporter.
DR   InterPro; IPR005829; Sug_transporter.
DR   InterPro; IPR003663; Sugar_transpt.
DR   Pfam; PF00083; sugar_tr; 1.
DR   PRINTS; PR00171; SUGRTRNSPORT.
DR   TIGRFAMs; TIGR00879; SP; 1.
DR   PROSITE; PS50850; MFS; 1.
DR   PROSITE; PS00216; SUGAR_TRANSPORT_1; 1.
DR   PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.
KW   Transmembrane; Sugar transport; Transport; Glycoprotein;
KW   Multigene family.
FT   DOMAIN        1      5       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      6     26       1 (POTENTIAL).
FT   DOMAIN       27     64       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     65     85       2 (POTENTIAL).
FT   DOMAIN       86     93       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     94    114       3 (POTENTIAL).
FT   DOMAIN      115    124       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    125    145       4 (POTENTIAL).
FT   DOMAIN      146    159       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    160    180       5 (POTENTIAL).
FT   DOMAIN      181    183       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    184    204       6 (POTENTIAL).
FT   DOMAIN      205    269       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    270    290       7 (POTENTIAL).
FT   DOMAIN      291    305       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    306    326       8 (POTENTIAL).
FT   DOMAIN      327    334       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    335    355       9 (POTENTIAL).
FT   DOMAIN      356    363       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    364    384       10 (POTENTIAL).
```

```
FT   DOMAIN       385    399       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     400    420       11 (POTENTIAL).
FT   DOMAIN       421    424       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     425    445       12 (POTENTIAL).
FT   DOMAIN       446    494       CYTOPLASMIC (POTENTIAL).
FT   SITE         277    279       DEFINES SUBSTRATE SPECIFICITY (BY
FT                                 SIMILARITY).
FT   CARBOHYD      43     43       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     358    358       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT     385    389       PWFIV -> SLVYC (IN REF. 2).
SQ   SEQUENCE   494 AA;  54019 MW;  EC42A3C648CAD23C CRC64;

  Query Match             23.6%;  Score 579.5;  DB 1;  Length 494;
  Best Local Similarity   31.2%;  Pred. No. 1.4e-33;
  Matches  157;  Conservative   96;  Mismatches  177;  Indels   73;  Gaps   14;

Qy         13 LGPPGGSAPRGRRVFLAAFAAALGPLSFGFALGY--SSPAI------PSLQRAAPPAPRL 64
              :|    :||    :  |   | :|   ||:  |   :  ||       :|:  : | |
Db          1 MGTTKVTAP----LIFAISVATIGSFQFGYNTGVINAPEAIIKDFLNYTLEERSEPPPSS 56

Qy         65 DDAAASWFGAVVTLGAAAGGVLG----GWLVDRAGRKLSLLLCSVPFVAGFAVI---TAA 117
                  : |  :|     : ||::|    |  |:| ||  |:|: ::  :||  ::     |
Db         57 VLLTSLWSLSVAIF--SVGGMIGSFSVGLFVNRFGRGNSMLIVNLLAIAGGCLMGFCKIA 114

Qy        118 QDVWMLLGGRLLTGLACGVASLVAPVYISEIAYPAVRGLLGSCVQLMVVVGILLAYLAGW 177
              : | ||: |||: || ||: :   |:| || :  |:||  |:  || :|:|||:| : |
Db        115 ESVEMLILGRLIIGLFCGLCTGFVPMYIGEISPTALRGAFGTLNQLGIVIGILVAQIFGL 174

Qy        178 VL----EWRWLAVLG--CVPPSLMLLLMCFMPETPRFLLTQHRRQE-AMAALRFLWGSEQ 230
               :    |  |  :||   :|  :    : | ||:|||||   : :| |   |: |||:|
Db        175 KVILGTEDLWPLLLGFTILPAIIQCAALPFCPESPRFLLINRKEEEKAKEILQRLWGTED 234

Qy        231 GWEDPPIGAEQSFH---------LALLRQPGIYKPFIIGVSLMAFQQLSGVNAVMFYAET 281
                :|     ::|           | | | |   :| || : |   |||||:||| :|:
Db        235 VAQDIQEMKDESMRMSQEKQVTVLELFRAPNYRQPIIISIMLQLSQQLSGINAVFYYSTG 294

Qy        282 IFEEAKFKDSSLASVVVGVIQVLFTAVAALIMDRAGRRLLLVLSGVVMVFSTSAFGAYFK 341
              ||::|  ::   |::  ||:  :|| |:  :::||||| | ::
Db        295 IFKDAGVQEPVYATIGAGVVNTIFTVVSVFLVERAGRRTLHLI----------------- 337

Qy        342 LTQGGPGNSSHVAISAPVSAQPVDASVGLAWLAVGSMCLFIAGFAVGWGPIPWLLMSEIF 401
                  | |  :  :|   :|    |    :::: :|:: :|:| | :| ||||| :::|:|
Db        338 ----GLGGMAFCSILMTISLLLKDNYSWMSFICIGAILVFVAFFEIGPGPIPWFIVAELF 393

Qy        402 PLHVKGVATGICVLTNWLMAFLVTKEFSSLMEVLRPYGAFWL-ASAFCIFSV------LF 454
                  :  |  :   :||   |||         :| |  ||:| |  | :|:|      :|
Db        394 GQGPRPAAMAVAGCSNWTSNFLV--------GLLFPSAAFYLGAYVFIVFTVFLVIFWVF 445

Qy        455 TLFCVPETKGKTLEQITAHFEGR 477
              | | ||||:|:| |:||  |||:
Db        446 TFFKVPETRGRTFEEITRAFEGQ 468


RESULT 12
GTR3_SHEEP
ID   GTR3_SHEEP     STANDARD;      PRT;   494 AA.
AC   P47843;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Solute carrier family 2, facilitated glucose transporter, member 3
DE   (Glucose transporter type 3, brain).
GN   SLC2A3 OR GLUT3 OR GLUT-3.
OS   Ovis aries (Sheep).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Caprinae; Ovis.
OX   NCBI_TaxID=9940;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Coopworth;
RX   MEDLINE=96109471; PubMed=8653093;
RA   Bennett B.L., Prosser C.G., Grigor M.R.;
RT   "Isolation of cDNAs and tissue specific expression of ovine glucose
RT   transporters.";
RL   Biochem. Mol. Biol. Int. 37:9-16(1995).
RN   [2]
RP   SEQUENCE OF 88-248 FROM N.A.
RC   TISSUE=Placenta;
RX   MEDLINE=97392487; PubMed=9250701;
RA   Currie M.J., Bassett N.S., Gluckman P.D.;
RT   "Ovine glucose transporter-1 and -3: cDNA partial sequences and
RT   developmental gene expression in the placenta.";
RL   Placenta 18:393-401(1997).
CC   -!- FUNCTION: Facilitative glucose transporter. Probably a neuronal
CC       glucose transporter.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- TISSUE SPECIFICITY: Highly expressed in brain.
CC   -!- SIMILARITY: Belongs to the sugar transporter family. Glucose
CC       transporter subfamily.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; L39214; AAC41629.1; -.
DR   EMBL; U89030; AAB49313.1; -.
DR   InterPro; IPR007114; MFS.
DR   InterPro; IPR005828; Sub_transporter.
DR   InterPro; IPR005829; Sug_transporter.
DR   InterPro; IPR003663; Sugar_transpt.
DR   Pfam; PF00083; sugar_tr; 1.
DR   PRINTS; PR00171; SUGRTRNSPORT.
DR   TIGRFAMs; TIGR00879; SP; 1.
DR   PROSITE; PS50850; MFS; 1.
DR   PROSITE; PS00216; SUGAR_TRANSPORT_1; 1.
DR   PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.
KW   Transmembrane; Sugar transport; Transport; Glycoprotein;
KW   Multigene family.
FT   DOMAIN         1      5       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM       6     26       1 (POTENTIAL).
FT   DOMAIN        27     64       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM      65     85       2 (POTENTIAL).
FT   DOMAIN        86     93       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      94    114       3 (POTENTIAL).
FT   DOMAIN       115    124       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     125    145       4 (POTENTIAL).
FT   DOMAIN       146    153       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     154    174       5 (POTENTIAL).
FT   DOMAIN       175    183       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     184    204       6 (POTENTIAL).
FT   DOMAIN       205    269       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     270    290       7 (POTENTIAL).
FT   DOMAIN       291    304       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     305    325       8 (POTENTIAL).
FT   DOMAIN       326    334       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     335    355       9 (POTENTIAL).
FT   DOMAIN       356    363       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     364    384       10 (POTENTIAL).
FT   DOMAIN       385    399       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     400    420       11 (POTENTIAL).
FT   DOMAIN       421    424       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     425    445       12 (POTENTIAL).
FT   DOMAIN       446    494       CYTOPLASMIC (POTENTIAL).
FT   SITE         277    279       DEFINES SUBSTRATE SPECIFICITY (BY
FT                                 SIMILARITY).
FT   CARBOHYD      43     43       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   494 AA;  54194 MW;  A89204D3EA74BFBA CRC64;

  Query Match             23.4%;  Score 575.5;  DB 1;  Length 494;
  Best Local Similarity   30.7%;  Pred. No. 2.6e-33;
  Matches  150;  Conservative   94;  Mismatches  179;  Indels   65;  Gaps   10;
```

```
Qy         26 VFLAAFAAALGPLSFGFALGYSSPAIPSL----------QRAAPPAPRLDDAAASWFGAV 75
              :  |   | :|   ||:  |  :     :          :   ||:  |  :  |   |:
Db         10 LIFAISIATIGSFQFGYNTGVINAPEAIIKDFLNYTLEERSETPPSSVLLTSLWSLSVAI 69

Qy         76 VTLGAAAGGVLGGWLVDRAGRKLSLLLCSVPFVAGFAVI---TAAQDVWMLLGGRLLTGL 132
               ::|   |    |  |:| ||: |:|: ::  :||  ::     |: | ||: |||: ||
Db         70 FSVGGMIGSFSVGLFVNRFGRRNSMLIVNLLAIAGGCLMGFCKIAESVEMLILGRLIIGL 129

Qy        133 ACGVASLVAPVYISEIAYPAVRGLLGSCVQLMVVVGILLAYLAGWVL----EWRWLAVLG 188
               ||: :   |:|| ||:  |:||  |:  || :|:|||:| : |  :    |  |  :||
Db        130 FCGLCTGFVPMYIGEISPTALRGAFGTLNQLGIVIGILVAQIFGLKVILGTEDLWPLLLG 189

Qy        189 --CVPPSLMLLLMCFMPETPRFLLTQHRRQE-AMAALRFLWGSEQGWEDPPIGAEQSFH- 244
                 :|  :    : | ||:|||||   : :| |   |: |||:|   :|     ::|
Db        190 FTILPAIIQCAALPFCPESPRFLLINRKEEEKAKEILQRLWGTEDVAQDIQEMKDESMRM 249

Qy        245 --------LALLRQPGIYKPFIIGVSLMAFQQLSGVNAVMFYAETIFEEAKFKDSSLASV 296
                      | | | |   :| || : |   |||||:||| :|:  ||::|  ::   |::
Db        250 SQEKQVTVLELFRAPNYRQPIIISIMLQLSQQLSGINAVFYYSTGIFKDAGVQEPVYATI 309

Qy        297 VVGVIQVLFTAVAALIMDRAGRRLLLVLSGVVMVFSTSAFGAYFKLTQGGPGNSSHVAIS 356
                ||:  :|| |:  :::||||| | ::                     | |  :  :|
Db        310 GAGVVNTIFTVVSVFLVERAGRRTLHLI---------------------GLGGMAFCSIL 348

Qy        357 APVSAQPVDASVGLAWLAVGSMCLFIAGFAVGWGPIPWLLMSEIFPLHVKGVATGICVLT 416
                :|    |    :::: :|:: :|:| | :| ||||| :::|:|    :  |  :   :
Db        349 MTISLLLKDNYSWMSFICIGAILVFVAFFEIGPGPIPWFIVAELFGQGPRPAAMAVAGCS 408

Qy        417 NWLMAFLVTKEFSSLMEVLRPYGAFWL-ASAFCIFSV------LFTLFCVPETKGKTLEQ 469
              ||   |||         :| |   |:| |  | :|:|      :|| | ||||:|:| |:
Db        409 NWTSNFLV--------GLLFPSATFYLGAYVFIVFTVFLVIFWVFTFFKVPETRGRTFEE 460

Qy        470 ITAHFEGR 477
              ||  |||:
Db        461 ITRAFEGQ 468


RESULT 13
GTR3_RAT
ID   GTR3_RAT       STANDARD;      PRT;   493 AA.
AC   Q07647; Q62729;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Solute carrier family 2, facilitated glucose transporter, member 3
DE   (Glucose transporter type 3, brain).
GN   SLC2A3 OR GLUT3 OR GLUT-3.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94063043; PubMed=8243635;
RA   Nagamatsu S., Sawa H., Kamada K., Nakamichi Y., Yoshimoto K.,
RA   Hoshino T.;
RT   "Neuron-specific glucose transporter (NSGT): CNS distribution of
RT   GLUT3 rat glucose transporter (RGT3) in rat central neurons.";
RL   FEBS Lett. 334:289-295(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Brain;
RX   MEDLINE=96042338; PubMed=7475896;
RA   Krishnan S.N., Haddad G.G.;
RT   "Cloning of glucose transporter-3 (GLUT3) cDNA from rat brain.";
RL   Life Sci. 56:1193-1197(1995).
CC   -!- FUNCTION: Facilitative glucose transporter. Probably a neuronal
CC       glucose transporter.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- TISSUE SPECIFICITY: Highly expressed in brain.
CC   -!- SIMILARITY: Belongs to the sugar transporter family. Glucose
CC       transporter subfamily.
```

```
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; D13962; BAA03065.1; -.
DR   EMBL; U17978; AAA62503.1; -.
DR   PIR; S38981; S38981.
DR   InterPro; IPR007114; MFS.
DR   InterPro; IPR005828; Sub_transporter.
DR   InterPro; IPR005829; Sug_transporter.
DR   InterPro; IPR003663; Sugar_transpt.
DR   Pfam; PF00083; sugar_tr; 1.
DR   PRINTS; PR00171; SUGRTRNSPORT.
DR   TIGRFAMs; TIGR00879; SP; 1.
DR   PROSITE; PS50850; MFS; 1.
DR   PROSITE; PS00216; SUGAR_TRANSPORT_1; 1.
DR   PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.
KW   Transmembrane; Sugar transport; Transport; Glycoprotein;
KW   Multigene family.
FT   DOMAIN        1     10       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     11     31       1 (POTENTIAL).
FT   DOMAIN       32     64       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     65     85       2 (POTENTIAL).
FT   DOMAIN       86     93       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     94    114       3 (POTENTIAL).
FT   DOMAIN      115    124       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    125    145       4 (POTENTIAL).
FT   DOMAIN      146    153       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    154    174       5 (POTENTIAL).
FT   DOMAIN      175    183       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    184    204       6 (POTENTIAL).
FT   DOMAIN      205    269       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    270    290       7 (POTENTIAL).
FT   DOMAIN      291    304       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    305    325       8 (POTENTIAL).
FT   DOMAIN      326    334       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    335    355       9 (POTENTIAL).
FT   DOMAIN      356    363       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    364    384       10 (POTENTIAL).
FT   DOMAIN      385    399       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    400    420       11 (POTENTIAL).
FT   DOMAIN      421    424       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    425    445       12 (POTENTIAL).
FT   DOMAIN      446    493       CYTOPLASMIC (POTENTIAL).
FT   SITE        277    279       DEFINES SUBSTRATE SPECIFICITY (BY
FT                                SIMILARITY).
FT   CARBOHYD     43     43       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT     55     55       R -> S (IN REF. 2).
FT   CONFLICT    447    447       S -> F (IN REF. 2).
FT   CONFLICT    470    470       H -> Q (IN REF. 2).
SQ   SEQUENCE   493 AA;  53580 MW;  745446B59BDF4399 CRC64;

  Query Match             23.4%;  Score 574.5;  DB 1;  Length 493;
  Best Local Similarity   31.1%;  Pred. No. 3e-33;
  Matches  156;  Conservative   93;  Mismatches  176;  Indels   77;  Gaps   15;

Qy         18 GSAPRGRRVFLAAFAAALGPLSFGFALGYSSPAIPSLQRAAPPA-----------PRLDD 66
              |:|     :  |   | :|   ||:  |  :         ||              ||:|
Db          2 GTAKVTPSLVFAVTVATIGSFQFGYNTGVIN---------APETIIKDFLNYTLEERLED 52

Qy         67 AAASWFGAVVTLGA------AAGGVLG----GWLVDRAGRKLSLLLCSVPFV-----AGF 111
                    | : || :      : ||::|    |  |:| ||: |:|| ::  :      ||
Db         53 LPRE--GLLTTLWSLCVAIFSVGGMIGSFSVGLFVNRFGRRNSMLLVNLIAILGGCLMGF 110

Qy        112 AVITAAQDVWMLLGGRLLTGLACGVASLVAPVYISEIAYPAVRGLLGSCVQLMVVVGILL 171
              | |  |: | ||: |||: |: ||: :   |:|| |::  |:||  |:  || :|||||:
Db        111 AKI--AESVEMLILGRLIIGIFCGLCTGFVPMYIGEVSPTALRGAFGTLNQLGIVVGILV 168
```

```
Qy     172 AYLAG--WVL--EWRWLAVLG--CVPPSLMLLLMCFMPETPRFLLTQHRRQ-EAMAALRF 224
           | : |  ::|  |  |  :||   :|  |    : | ||:|||||   : : :|   |:
Db     169 AQVFGLDFILGSEELWPGLLGLTIIPAILQSAALPFCPESPRFLLINRKEEDQATEILQR 228

Qy     225 LWGSE------QGWEDPPIGAEQSFH---LALLRQPGIYKPFIIGVSLMAFQQLSGVNAV 275
           |||:       |  :|  |   |      | | : |  ::| :| | |   || ||:|||
Db     229 LWGTPDVIQEIQEMKDESIRMSQEKQVTVLELFKSPSYFQPLLISVVLQLSQQFSGINAV 288

Qy     276 MFYAETIFEEAKFKDSSLASVVVGVIQVLFTAVAALIMDRAGRRLLLVLSGVVMVFSTSA 335
            :|:  ||::|  ::   |::  ||:  :|| |:  :::||||| | ::
Db     289 FYYSTGIFQDAGVQEPIYATIGAGVVNTIFTVVSLFLVERAGRRTLHMI----------- 337

Qy     336 FGAYFKLTQGGPGNSSHVAISAPVSAQPVDASVGLAWLAVGSMCLFIAGFAVGWGPIPWL 395
                     | |  :  ::   :|    |    :::: : :: :::| | :| |||||
Db     338 ----------GLGGMAVCSVFMTISLLLKDEYEAMSFVCIVAILVYVAFFEIGPGPIPWF 387

Qy     396 LMSEIFPLHVKGVATGICVLTNWLMAFLVTKEFSSLMEVLRPYGAFWLASAFCIFSVLFT 455
           :::|:|    :  |  :   :||   |||   | |    |  |  | : :|| :| ::||
Db     388 IVAELFSQGPRPAAMAVAGCSNWTSNFLVGMFFPSAAAYLGAY-VFIIFAAFLVFFLIFT 446

Qy     456 LFCVPETKGKTLEQITAHFEGR 477
            | ||||||:| | ||  |||:
Db     447 SFKVPETKGRTFEDITRAFEGQ 468

RESULT 14
GTR3_HUMAN
ID   GTR3_HUMAN     STANDARD;      PRT;   496 AA.
AC   P11169; Q9UG15;
DT   01-JUL-1989 (Rel. 11, Created)
DT   01-JUL-1989 (Rel. 11, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Solute carrier family 2, facilitated glucose transporter, member 3
DE   (Glucose transporter type 3, brain).
GN   SLC2A3 OR GLUT3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89008414; PubMed=3170580;
RA   Kayano T., Fukumoto H., Eddy R.L., Fan Y.-S., Byers M.G., Shows T.B.,
RA   Bell G.I.;
RT   "Evidence for a family of human glucose transporter-like proteins.
RT   Sequence and gene localization of a protein expressed in fetal
RT   skeletal muscle and other tissues.";
RL   J. Biol. Chem. 263:15245-15248(1988).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Stuart C.A., Wen K.G., Acosta M., Wood T.G.;
RT   "Resistance and expression of glucose transporters in human skeletal
RT   muscle.";
RL   Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 288-496 FROM N.A.
RC   TISSUE=Fetal brain;
RA   Blum H., Bauersachs S., Mewes H.-W., Gassenhuber J., Wiemann S.;
RL   Submitted (AUG-1999) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 302-453 FROM N.A.
RC   TISSUE=Articular cartilage;
RA   Neama G., Richardson S., Bell S., Carter S., Mobasheri A.;
RT   "Molecular characterization and cloning of glucose transporters in
RT   human articular chondrocytes.";
RL   Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SUBSTRATE SPECIFICITY SITE.
RX   MEDLINE=98138446; PubMed=9477959;
RA   Seatter M.J., de la Rue S.A., Porter L.M., Gould G.W.;
RT   "QLS motif in transmembrane helix VII of the glucose transporter
RT   family interacts with the C-1 (position) of D-glucose and is involved
RT   in substrate selection at the exofacial binding site.";
RL   Biochemistry 37:1322-1326(1998).
CC   -!- FUNCTION: Facilitative glucose transporter. Probably a neuronal
CC       glucose transporter.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- TISSUE SPECIFICITY: Highly expressed in brain.
CC   -!- SIMILARITY: Belongs to the sugar transporter family. Glucose
CC       transporter subfamily.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M20681; AAB61083.1; -.
DR   EMBL; AF274892; AAF82116.1; -.
DR   EMBL; AF274889; AAF82116.1; JOINED.
DR   EMBL; AF274890; AAF82116.1; JOINED.
DR   EMBL; AF274891; AAF82116.1; JOINED.
DR   EMBL; AL110298; CAB53739.1; -.
DR   EMBL; AY034634; AAK56796.1; -.
DR   PIR; A31986; A31986.
DR   PIR; T14798; T14798.
DR   Genew; HGNC:11007; SLC2A3.
DR   GK; P11169; -.
DR   MIM; 138170; -.
DR   GO; GO:0005355; F:glucose transporter activity; TAS.
DR   GO; GO:0015758; P:glucose transport; TAS.
DR   InterPro; IPR007114; MFS.
DR   InterPro; IPR005828; Sub_transporter.
DR   InterPro; IPR005829; Sug_transporter.
DR   InterPro; IPR003663; Sugar_transpt.
DR   Pfam; PF00083; sugar_tr; 1.
DR   PRINTS; PR00171; SUGRTRNSPORT.
DR   TIGRFAMs; TIGR00879; SP; 1.
DR   PROSITE; PS50850; MFS; 1.
DR   PROSITE; PS00216; SUGAR_TRANSPORT_1; 1.
DR   PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.
KW   Transmembrane; Sugar transport; Transport; Glycoprotein;
KW   Multigene family.
FT   DOMAIN        1      5       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      6     26       1 (POTENTIAL).
FT   DOMAIN       27     64       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     65     85       2 (POTENTIAL).
FT   DOMAIN       86     93       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     94    114       3 (POTENTIAL).
FT   DOMAIN      115    124       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    125    145       4 (POTENTIAL).
FT   DOMAIN      146    153       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    154    174       5 (POTENTIAL).
FT   DOMAIN      175    183       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    184    204       6 (POTENTIAL).
FT   DOMAIN      205    269       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    270    290       7 (POTENTIAL).
FT   DOMAIN      291    304       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    305    325       8 (POTENTIAL).
FT   DOMAIN      326    334       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    335    355       9 (POTENTIAL).
FT   DOMAIN      356    368       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    369    389       10 (POTENTIAL).
FT   DOMAIN      390    399       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    400    420       11 (POTENTIAL).
FT   DOMAIN      421    427       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    428    448       12 (POTENTIAL).
FT   DOMAIN      449    496       CYTOPLASMIC (POTENTIAL).
FT   SITE        277    279       DEFINES SUBSTRATE SPECIFICITY (BY
FT                                SIMILARITY).
FT   CARBOHYD     43     43       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT    303    303       E -> Q (IN REF. 3).
FT   CONFLICT    310    310       G -> S (IN REF. 3).
FT   CONFLICT    320    321       VV -> LL (IN REF. 3).
```

```
FT   CONFLICT    357    358       DN -> NH (IN REF. 3).
FT   CONFLICT    376    376       F -> C (IN REF. 3).
FT   CONFLICT    425    425       H -> Y (IN REF. 3).
FT   CONFLICT    432    432       F -> L (IN REF. 3).
FT   CONFLICT    482    482       E -> G (IN REF. 3).
SQ   SEQUENCE   496 AA;  53924 MW;  F601CD6892F16516 CRC64;

  Query Match             23.2%;  Score 570.5;  DB 1;  Length 496;
  Best Local Similarity   31.6%;  Pred. No. 5.8e-33;
  Matches  152;  Conservative   88;  Mismatches  190;  Indels   51;  Gaps    9;

Qy         26 VFLAAFAAALGPLSFGFALGYSSPAIPSLQR----------AAPPAPRLDDAAASWFGAV 75
              :  |   | :|   ||:  |  :     ::            |||:  |  :  |   |:
Db         10 LIFAITVATIGSFQFGYNTGVINAPEKIIKEFINKTLTDKGNAPPSEVLLTSLWSLSVAI 69

Qy         76 VTLGAAAGGVLGGWLVDRAGRKLSLLLCSVPFVAG---FAVITAAQDVWMLLGGRLLTGL 132
               ::|   |    |  |:| ||: |:|: ::  | |     :   |: | ||: |||: ||
Db         70 FSVGGMIGSFSVGLFVNRFGRRNSMLIVNLLAVTGGCFMGLCKVAKSVEMLILGRLVIGL 129

Qy        133 ACGVASLVAPVYISEIAYPAVRGLLGSCVQLMVVVGILLAYLAG--WVL--EWRWLAVLG 188
               ||: :   |:|| ||:  |:||  |:  || :|||||:| : |  ::|  |  |  :||
Db        130 FCGLCTGFVPMYIGEISPTALRGAFGTLNQLGIVVGILVAQIFGLEFILGSEELWPLLLG 189

Qy        189 --CVPPSLMLLLMCFMPETPRFLLTQHRRQE-AMAALRFLWGSEQGWEDPPIGAEQSFH- 244
                 :|  |    :  | ||:|||||   : :| |   |: |||::   :|     ::|
Db        190 FTILPAILQSAALPFCPESPRFLLINRKEEENAKQILQRLWGTQDVSQDIQEMKDESARM 249

Qy        245 --------LALLRQPGIYKPFIIGVSLMAFQQLSGVNAVMFYAETIFEEAKFKDSSLASV 296
                      | | |     :| || : |   |||||:||| :|:  ||::|  ::   |::
Db        250 SQEKQVTVLELFRVSSYRQPIIISIVLQLSQQLSGINAVFYYSTGIFKDAGVQEPIYATI 309

Qy        297 VVGVIQVLFTAVAALIMDRAGRRLLLVLSGVVMVFSTSAFGAYFKLTQGGPGNSSHVAIS 356
                ||:  :|| |:  :::||||| | ::                     | |  :  :
Db        310 GAGVVNTIFTVVSLFLVERAGRRTLHMI---------------------GLGGMAFCSTL 348

Qy        357 APVSAQPVDASVGLAWLAVGSMCLFIAGFAVGWGPIPWLLMSEIFPLHVKGVATGICVLT 416
                ||    |   |:::: :|:: :|:| | :| ||||| :::|:|    :  |  :   :
Db        349 MTVSLLLKDNYNGMSFVCIGAILVFVAFFEIGPGPIPWFIVAELFSQGPRPAAMAVAGCS 408

Qy        417 NWLMAFLVTKEFSSLMEVLRPYGAFWLASAFCIFSVLFTLFCVPETKGKTLEQITAHFEG 476
              ||   |||   | |    |  |  | : : | |  : || | ||||:|:| | ||  |||
Db        409 NWTSNFLVGLLFPSAAHYLGAY-VFIIFTGFLITFLAFTFFKVPETRGRTFEDITRAFEG 467

Qy        477 R 477
              :
Db        468 Q 468


RESULT 15
GTR3_MOUSE
ID   GTR3_MOUSE     STANDARD;      PRT;   493 AA.
AC   P32037;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Solute carrier family 2, facilitated glucose transporter, member 3
DE   (Glucose transporter type 3, brain).
GN   SLC2A3 OR GLUT3 OR GLUT-3.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=92112695; PubMed=1730609;
RA   Nagamatsu S., Kornhauser J.M., Seino S., Mayo K.E., Steiner D.F.,
RA   Bell G.I.;
RT   "Glucose transporter expression in brain. cDNA sequence of mouse
RT   GLUT3, the brain facilitative glucose transporter isoform, and
RT   identification of sites of expression by in situ hybridization.";
RL   J. Biol. Chem. 267:467-472(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Takeda J., Minokoshi Y., Yasuda K., Kayano T., Graeme B.I.;
RL   Submitted (JUL-1994) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Retina;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC   -!- FUNCTION: Facilitative glucose transporter. Probably a neuronal
CC       glucose transporter.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- TISSUE SPECIFICITY: Highly expressed in brain.
CC   -!- SIMILARITY: Belongs to the sugar transporter family. Glucose
CC       transporter subfamily.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M75135; AAA37704.1; -.
DR   EMBL; X61093; CAA43406.1; -.
DR   EMBL; U11853; AAB60666.1; -.
DR   EMBL; U11844; AAB60666.1; JOINED.
DR   EMBL; U11845; AAB60666.1; JOINED.
DR   EMBL; U11846; AAB60666.1; JOINED.
DR   EMBL; U11848; AAB60666.1; JOINED.
DR   EMBL; U11849; AAB60666.1; JOINED.
DR   EMBL; U11850; AAB60666.1; JOINED.
DR   EMBL; U11851; AAB60666.1; JOINED.
DR   EMBL; U11852; AAB60666.1; JOINED.
DR   EMBL; BC034122; AAH34122.1; -.
DR   EMBL; BC058811; AAH58811.1; -.
DR   PIR; A41751; A41751.
DR   MGD; MGI:95757; Slc2a3.
DR   InterPro; IPR007114; MFS.
DR   InterPro; IPR005828; Sub_transporter.
DR   InterPro; IPR005829; Sug_transporter.
DR   InterPro; IPR003663; Sugar_transpt.
DR   Pfam; PF00083; sugar_tr; 1.
DR   PRINTS; PR00171; SUGRTRNSPORT.
DR   TIGRFAMs; TIGR00879; SP; 1.
DR   PROSITE; PS50850; MFS; 1.
DR   PROSITE; PS00216; SUGAR_TRANSPORT_1; 1.
DR   PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.
KW   Transmembrane; Sugar transport; Transport; Glycoprotein;
KW   Multigene family.
FT   DOMAIN        1     10       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     11     31       1 (POTENTIAL).
FT   DOMAIN       32     64       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     65     85       2 (POTENTIAL).
```

```
FT   DOMAIN       86     93       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     94    114       3 (POTENTIAL).
FT   DOMAIN      115    124       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    125    145       4 (POTENTIAL).
FT   DOMAIN      146    153       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    154    174       5 (POTENTIAL).
FT   DOMAIN      175    183       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    184    204       6 (POTENTIAL).
FT   DOMAIN      205    269       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    270    290       7 (POTENTIAL).
FT   DOMAIN      291    304       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    305    325       8 (POTENTIAL).
FT   DOMAIN      326    334       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    335    355       9 (POTENTIAL).
FT   DOMAIN      356    363       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    364    384       10 (POTENTIAL).
FT   DOMAIN      385    399       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    400    420       11 (POTENTIAL).
FT   DOMAIN      421    427       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    428    448       12 (POTENTIAL).
FT   DOMAIN      449    493       CYTOPLASMIC (POTENTIAL).
FT   SITE        277    279       DEFINES SUBSTRATE SPECIFICITY (BY
FT                                SIMILARITY).
FT   CARBOHYD     43     43       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   493 AA;  53478 MW;  9090B8DCB8780082 CRC64;

  Query Match             23.2%;  Score 569.5;  DB 1;  Length 493;
  Best Local Similarity   31.5%;  Pred. No. 6.8e-33;
  Matches  154;  Conservative   86;  Mismatches  176;  Indels   73;  Gaps   13;

Qy          29 AAFAAALGPLSFGFALGYSSPAIPSLQRAAPPA-----------PRLDD--------AAA 69
               |   | :|   ||:  |  :         ||              ||:|        |
Db          13 AVTVATIGSFQFGYNTGVIN---------APETILKDFLNYTLEERLEDLPSEGLLTALW 63

Qy          70 SWFGAVVTLGAAAGGVLGGWLVDRAGRKLSLLLCSVPFV-----AGFAVITAAQDVWMLL 124
               |   |: ::|   |    |  |:| ||: |:|| ::  :      ||| |  |: | ||:
Db          64 SLCVAIFSVGGMIGSFSVGLFVNRFGRRNSMLLVNLLAIIAGCLMGFAKI--AESVEMLI 121

Qy         125 GGRLLTGLACGVASLVAPVYISEIAYPAVRGLLGSCVQLMVVVGILLAYLAG--WVL--E 180
                |||| |: ||: :   |:|| |::  |:||  |:  || :|||||:| : |  ::|  |
Db         122 LGRLLIGIFCGLCTGFVPMYIGEVSPTALRGAFGTLNQLGIVVGILVAQIFGLDFILGSE 181

Qy         181 WRWLAVLG--CVPPSLMLLLMCFMPETPRFLLTQHRRQ-EAMAALRFLWGSE------QG 231
                 |  :||   :|  |    : | ||:|||||   : : :|   |: |||:       |
Db         182 ELWPGLLGLTIIPAILQSAALPFCPESPRFLLINKKEEDQATEILQRLWGTSDVVQEIQE 241

Qy         232 WEDPPIGAEQSFH---LALLRQPGIYKPFIIGVSLMAFQQLSGVNAVMFYAETIFEEAKF 288
                :|  :   |      | | | |   :| :| : |   |||||:||| :|:  ||::|
Db         242 MKDESVRMSQEKQVTVLELFRSPNYVQPLLISIVLQLSQQLSGINAVFYYSTGIFKDAGV 301

Qy         289 KDSSLASVVVGVIQVLFTAVAALIMDRAGRRLLLVLSGVVMVFSTSAFGAYFKLTQGGPG 348
               ::   |::  ||:  :|| |:  :::||||| | ::                     | |
Db         302 QEPIYATIGAGVVNTIFTVVSLFLVERAGRRTLHMI---------------------GLG 340

Qy         349 NSSHVAISAPVSAQPVDASVGLAWLAVGSMCLFIAGFAVGWGPIPWLLMSEIFPLHVKGV 408
                 :  ::   :|    |    :::: : :: :::| | :| ||||| :::|:|    :
Db         341 GMAVCSVFMTISLLLKDDYEAMSFVCIVAILIYVAFFEIGPGPIPWFIVAELFSQGPRPA 400

Qy         409 ATGICVLTNWLMAFLVTKEFSSLMEVLRPYGAFWLASAFCIFSVLFTLFCVPETKGKTLE 468
               |  :    ||   |||   | |    |  |  | : :|| || ::|| | ||||||:| |
Db         401 AIAVAGCCNWTSNFLVGMLFPSAAAYLGAY-VFIIFAAFLIFFLIFTFFKVPETKGRTFE 459

Qy         469 QITAHFEGR 477
                |   |||:
Db         460 DIARAFEGQ 468

Search completed: September 27, 2004, 19:27:12
Job time : 28 secs
```

```
                    GenCore version 5.1.6
           Copyright (c) 1993 - 2004  Compugen Ltd.


OM protein - nucleic search, using frame_plus_p2n model

Run on:         September 28, 2004, 00:53:58 ; Search time 4589 Seconds
                                           (without alignments)
                                           4505.258 Million cell updates/sec

Title:          US-09-886-954A-1
Perfect score:  2457
Sequence:       1 MTPEDPEETQPLLGPPGGSA..........CVPETKGKTLEQITAHFEGR 477

Scoring table:  BLOSUM62
                Xgapop 10.0 , Xgapext  0.5
                Ygapop 10.0 , Ygapext  0.5
                Fgapop  6.0 , Fgapext  7.0
                Delop   6.0 , Delext   7.0

Searched:       3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters:      6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US09886954/runat_27092004_170631_13388/app_query.fasta_1.647
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09886954_@CGN_1_1_5600_@runat_27092004_170631_13388 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :      GenEmbl:*
                1:  gb_ba:*
                2:  gb_htg:*
                3:  gb_in:*
                4:  gb_om:*
                5:  gb_ov:*
                6:  gb_pat:*
                7:  gb_ph:*
                8:  gb_pl:*
                9:  gb_pr:*
                10:  gb_ro:*
                11:  gb_sts:*
                12:  gb_sy:*
                13:  gb_un:*
                14:  gb_vi:*
                15:  em_ba:*
                16:  em_fun:*
                17:  em_hum:*
                18:  em_in:*
                19:  em_mu:*
                20:  em_om:*
                21:  em_or:*
                22:  em_ov:*
                23:  em_pat:*
                24:  em_ph:*
                25:  em_pl:*
                26:  em_ro:*
                27:  em_sts:*
                28:  em_un:*
                29:  em_vi:*
                30:  em_htg_hum:*
                31:  em_htg_inv:*
                32:  em_htg_other:*
                33:  em_htg_mus:*
                34:  em_htg_pln:*
                35:  em_htg_rod:*
                36:  em_htg_mam:*
                37:  em_htg_vrt:*
                38:  em_sy:*
                39:  em_htgo_hum:*
                40:  em_htgo_mus:*
                41:  em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                              SUMMARIES
                      %
Result              Query
  No.      Score    Match Length DB  ID                      Description
---------------------------------------------------------------------------------
     1      2457    100.0   2080  6  AX179740                AX179740 Sequence
     2      2457    100.0   2145  9  BC019043                BC019043 Homo sapi
     3      2448     99.6   1873  9  HSA245937               AJ245937 Homo sapi
     4      2448     99.6   2217  6  AX076667                AX076667 Sequence
     5      2445     99.5   1508  9  HSA17801                Y17801 Homo sapien
     6    2162.5     88.0   2087  6  AX076669                AX076669 Sequence
     7    2162.5     88.0   2087 10  RNO245935                AJ245935 Rattus no
     8      2159     87.9   1490 10  MMU17802                 Y17802 Mus musculu
     9      2156     87.7   1843 10  AF232061                 AF232061 Mus muscu
    10      2150     87.5   2072  6  AX076671                AX076671 Sequence
    11      2150     87.5   2072 10  MMU245936                AJ245936 Mus muscu
    12      2146     87.3   2189 10  AB033418                 AB033418 Rattus no
    13    1554.5     63.3   1461  6  AX191507                AX191507 Sequence
    14    1537.5     62.6   1012  4  AF321324                AF321324 Bos tauru
    15    1411.5     57.4   1929  5  BC049409                BC049409 Danio rer
    16      1407     57.3   1449  5  AB083371                AB083371 Gallus ga
    17      1268     51.6    789  6  AX191497                AX191497 Sequence
    18      1167     47.5   1493 10  BC021758                 BC021758 Mus muscu
    19    1125.5     45.8    738  4  AF495799                AF495799 Ovis arie
    20     972.5     39.6   2011  6  AX076679                AX076679 Sequence
    21     958.5     39.0   1580  9  HSA17803                Y17803 Homo sapien
    22     958.5     39.0   2356  6  AX677165                AX677165 Sequence
    23     958.5     39.0   2356  9  BC013740                BC013740 Homo sapi
    24     958.5     39.0   2390  6  BD127367                BD127367 Primer fo
    25     958.5     39.0   2390  9  AK074927                AK074927 Homo sapi
    26     958.5     39.0   2487  9  HSA011372               AJ011372 Homo sapi
    27       949     38.6   1541  6  AX076677                AX076677 Sequence
    28     910.5     37.1   2364 10  BC058210                 BC058210 Mus muscu
    29     904.5     36.8   2171  6  BD127313                BD127313 Primer fo
    30     904.5     36.8   2171  9  AK074836                AK074836 Homo sapi
    31     891.5     36.3   9582 10  MMU413951                AJ413951 Mus muscu
    32     891.5     36.3 198006  2  AL731852                AL731852 Mus muscu
    33     870.5     35.4 225370  9  AL445222                AL445222 Human DNA
    34     867.5     35.3   3114  9  AK096630                AK096630 Homo sapi
c   35     824.5     33.6 250940  2  AC096489                AC096489 Rattus no
c   36       807     32.8 241623  2  AC103008                AC103008 Rattus no
    37     804.5     32.7   1673  3  AY122079                AY122079 Drosophil
    38     789.5     32.1   3160  3  BT003466                BT003466 Drosophil
    39     788.5     32.1   2093  3  BT010286                BT010286 Drosophil
    40       788     32.1   2937  3  AY119564                AY119564 Drosophil
    41       736     30.0   1557  8  AY093274                AY093274 Arabidops
    42       736     30.0   1866  8  AY059848                AY059848 Arabidops
    43       734     29.9   1464  8  AY124845                AY124845 Arabidops
    44       734     29.9   1587  3  AY058435                AY058435 Drosophil
    45       734     29.9   1695  8  AF412060                AF412060 Arabidops


                              ALIGNMENTS

RESULT 1
```

```
AX179740
LOCUS       AX179740                 2080 bp    DNA     linear   PAT 06-AUG-2001
DEFINITION  Sequence 28 from Patent WO0146258.
ACCESSION   AX179740
VERSION     AX179740.1  GI:15132104
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Baughn,M.R., Burford,N., Au-Young,J., Lu,D.A., Yang,J., Reddy,R.,
            Lal,P., Hillman,J.L., Azimzai,Y., Yue,H., Nguyen,D.B., Yao,M.G.,
            Gandhi,A.R., Tang,Y.T. and Khan,F.A.
  TITLE     Transporters and ion channels
  JOURNAL   Patent: WO 0146258-A 28 28-JUN-2001;
            Incyte Genomics, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .2080
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
                     /note="Incyte ID No: 1416107CB1"
ORIGIN

Alignment Scores:
Pred. No.:              2.44e-167      Length:         2080
Score:                  2457.00        Matches:        477
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     6              Gaps:           0

US-09-886-954A-1 (1-477) x AX179740 (1-2080)

Qy          1 MetThrProGluAspProGluGluThrGlnProLeuLeuGlyProProGlyGlySerAla 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         46 ATGACGCCCGAGGACCCAGAGGAAACCCAGCCGCTTCTGGGGCCTCCTGGCGGCAGCGCG 105

Qy         21 ProArgGlyArgArgValPheLeuAlaAlaPheAlaAlaAlaLeuGlyProLeuSerPhe 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        106 CCCCGCGGCCGCCGCGTCTTCCTCGCCGCCTTCGCCGCTGCCCTGGGCCCACTCAGCTTC 165

Qy         41 GlyPheAlaLeuGlyTyrSerSerProAlaIleProSerLeuGlnArgAlaAlaProPro 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        166 GGCTTCGCGCTCGGCTACAGCTCCCCGGCCATCCCTAGCCTGCAGCGCGCCGCGCCCCCG 225

Qy         61 AlaProArgLeuAspAspAlaAlaAlaSerTrpPheGlyAlaValValThrLeuGlyAla 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        226 GCCCCGCGCCTGGACGACGCCGCCGCCTCCTGGTTCGGGGCTGTCGTGACCCTGGGTGCC 285

Qy         81 AlaAlaGlyGlyValLeuGlyGlyTrpLeuValAspArgAlaGlyArgLysLeuSerLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        286 GCGGCGGGGGGAGTGCTGGGCGGCTGGCTGGTGGACCGCGCCGGGCGCAAGCTGAGCCTC 345

Qy        101 LeuLeuCysSerValProPheValAlaGlyPheAlaValIleThrAlaAlaGlnAspVal 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        346 TTGCTGTGCTCCGTGCCCTTCGTGGCCGGCTTTGCCGTCATCACCGCGGCCCAGGACGTG 405

Qy        121 TrpMetLeuLeuGlyGlyArgLeuLeuThrGlyLeuAlaCysGlyValAlaSerLeuVal 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        406 TGGATGCTGCTGGGGGGCCGCCTCCTCACCGGCCTGGCCTGCGGTGTTGCCTCCCTAGTG 465

Qy        141 AlaProValTyrIleSerGluIleAlaTyrProAlaValArgGlyLeuLeuGlySerCys 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        466 GCCCCGGTCTACATCTCCGAAATCGCCTACCCAGCAGTCCGGGGGTTGCTCGGCTCCTGT 525

Qy        161 ValGlnLeuMetValValValGlyIleLeuLeuAlaTyrLeuAlaGlyTrpValLeuGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        526 GTGCAGCTAATGGTCGTCGTCGGCATCCTCCTGGCCTACCTGGCAGGCTGGGTGCTGGAG 585

Qy        181 TrpArgTrpLeuAlaValLeuGlyCysValProProSerLeuMetLeuLeuLeuMetCys 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        586 TGGCGCTGGCTGGCTGTGCTGGGCTGCGTGCCCCCCTCCCTCATGCTGCTTCTCATGTGC 645

Qy        201 PheMetProGluThrProArgPheLeuLeuThrGlnHisArgArgGlnGluAlaMetAla 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        646 TTCATGCCCGAGACCCCGCGCTTCCTGCTGACTCAGCACAGGCGCCAGGAGGCCATGGCC 705

Qy        221 AlaLeuArgPheLeuTrpGlySerGluGlnGlyTrpGluAspProProIleGlyAlaGlu 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        706 GCCCTGCGGTTCCTGTGGGGCTCCGAGCAGGGCTGGGAAGACCCCCCCATCGGGGCTGAG 765

Qy        241 GlnSerPheHisLeuAlaLeuLeuArgGlnProGlyIleTyrLysProPheIleIleGly 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        766 CAGAGCTTTCACCTGGCCCTGCTGCGGCAGCCCGGCATCTACAAGCCCTTCATCATCGGC 825

Qy        261 ValSerLeuMetAlaPheGlnGlnLeuSerGlyValAsnAlaValMetPheTyrAlaGlu 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        826 GTCTCCCTGATGGCCTTCCAGCAGCTGTCGGGGGTCAACGCCGTCATGTTCTATGCAGAG 885

Qy        281 ThrIlePheGluGluAlaLysPheLysAspSerSerLeuAlaSerValValValGlyVal 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        886 ACCATCTTTGAAGAGGCCAAGTTCAAGGACAGCAGCCTGGCCTCGGTCGTCGTGGGTGTC 945

Qy        301 IleGlnValLeuPheThrAlaValAlaAlaLeuIleMetAspArgAlaGlyArgArgLeu 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        946 ATCCAGGTGCTGTTCACAGCTGTGGCGGCTCTCATCATGGACAGAGCAGGGCGGAGGCTG 1005

Qy        321 LeuLeuValLeuSerGlyValValMetValPheSerThrSerAlaPheGlyAlaTyrPhe 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1006 CTCCTGGTCTTGTCAGGTGTGGTCATGGTGTTCAGCACGAGTGCCTTCGGCGCCTACTTC 1065

Qy        341 LysLeuThrGlnGlyGlyProGlyAsnSerSerHisValAlaIleSerAlaProValSer 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1066 AAGCTGACCCAGGGTGGCCCTGGCAACTCCTCGCACGTGGCCATCTCGGCGCCTGTCTCT 1125

Qy        361 AlaGlnProValAspAlaSerValGlyLeuAlaTrpLeuAlaValGlySerMetCysLeu 380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1126 GCACAGCCTGTTGATGCCAGCGTGGGGCTGGCCTGGCTGGCCGTGGGCAGCATGTGCCTC 1185

Qy        381 PheIleAlaGlyPheAlaValGlyTrpGlyProIleProTrpLeuLeuMetSerGluIle 400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1186 TTCATCGCCGGCTTTGCGGTGGGCTGGGGGCCCATCCCCTGGCTCCTCATGTCAGAGATC 1245

Qy        401 PheProLeuHisValLysGlyValAlaThrGlyIleCysValLeuThrAsnTrpLeuMet 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1246 TTCCCTCTGCATGTCAAGGGCGTGGCGACAGGCATCTGCGTCCTCACCAACTGGCTCATG 1305

Qy        421 AlaPheLeuValThrLysGluPheSerSerLeuMetGluValLeuArgProTyrGlyAla 440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1306 GCCTTTCTCGTGACCAAGGAGTTCAGCAGCCTCATGGAGGTCCTCAGGCCCTATGGAGCC 1365

Qy        441 PheTrpLeuAlaSerAlaPheCysIlePheSerValLeuPheThrLeuPheCysValPro 460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1366 TTCTGGCTTGCCTCCGCTTTCTGCATCTTCAGTGTCCTTTTCACTTTGTTCTGTGTCCCT 1425

Qy        461 GluThrLysGlyLysThrLeuGluGlnIleThrAlaHisPheGluGlyArg 477
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1426 GAAACTAAAGGAAAGACTCTGGAACAAATCACAGCCCATTTTGAGGGGCGA 1476

RESULT 2
BC019043
LOCUS       BC019043                 2145 bp    mRNA    linear   PRI 19-DEC-2003
DEFINITION  Homo sapiens solute carrier family 2, (facilitated glucose
            transporter) member 8, mRNA (cDNA clone MGC:20634 IMAGE:4641145),
            complete cds.
ACCESSION   BC019043
VERSION     BC019043.2  GI:40226554
KEYWORDS    MGC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2145)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
```

```
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
REFERENCE   2  (bases 1 to 2145)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-DEC-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     On Dec 19, 2003 this sequence version replaced gi:17512129.
            Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC/DCTD/DTP
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: National Institutes of Health Intramural
            Sequencing Center (NISC),
            Gaithersburg, Maryland;
            Web site: http://www.nisc.nih.gov/
            Contact: nisc_mgc@nhgri.nih.gov
            Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
            Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
            Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
            Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R.,
            Maduro,Q.L., Masiello,C., Maskeri,B., Mastrian,S.D.,McCloskey,J.C.,
            McDowell,J., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W.,
            Tsurgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
            Young,A., Zhang,L.-H. and Green,E.D.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 30 Row: n Column: 3
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 21361448.
FEATURES             Location/Qualifiers
     source          1. .2145
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="MGC:20634 IMAGE:4641145"
                     /tissue_type="Skin, melanotic melanoma."
                     /clone_lib="NIH_MGC_20"
                     /lab_host="DH10B-R"
                     /note="Vector: pOTB7"
     gene            1. .2145
                     /gene="SLC2A8"
                     /note="synonyms: GLUTX1, GLUT8"
                     /db_xref="LocusID:29988"
                     /db_xref="MIM:605245"
     CDS             46. .1479
                     /gene="SLC2A8"
                     /codon_start=1
                     /product="solute carrier family 2, (facilitated glucose
                     transporter) member 8"
```

```
                     /protein_id="AAH19043.1"
                     /db_xref="GI:17512130"
                     /db_xref="LocusID:29988"
                     /db_xref="MIM:605245"
                     /translation="MTPEDPEETQPLLGPPGGSAPRGRRVFLAAFAAALGPLSFGFAL
                     GYSSPAIPSLQRAAPPAPRLDDAAASWFGAVVTLGAAAGGVLGGWLVDRAGRKLSLLL
                     CSVPFVAGFAVITAAQDVWMLLGGRLLTGLACGVASLVAPVYISEIAYPAVRGLLGSC
                     VQLMVVVGILLAYLAGWVLEWRWLAVLGCVPPSLMLLLMCFMPETPRFLLTQHRRQEA
                     MAALRFLWGSEQGWEDPPIGAEQSFHLALLRQPGIYKPFIIGVSLMAFQQLSGVNAVM
                     FYAETIFEEAKFKDSSLASVVVGVIQVLFTAVAALIMDRAGRRLLLVLSGVVMVFSTS
                     AFGAYFKLTQGGPGNSSHVAISAPVSAQPVDASVGLAWLAVGSMCLFIAGFAVGWGPI
                     PWLLMSEIFPLHVKGVATGICVLTNWLMAFLVTKEFSSLMEVLRPYGAFWLASAFCIF
                     SVLFTLFCVPETKGKTLEQITAHFEGR"
     misc_feature    319. .1467
                     /gene="SLC2A8"
                     /note="sugar_tr; Region: Sugar (and other) transporter"
                     /db_xref="CDD:pfam00083"
ORIGIN

Alignment Scores:
Pred. No.:               2.53e-167      Length:         2145
Score:                   2457.00        Matches:        477
Percent Similarity:      100.00%        Conservative:   0
Best Local Similarity:   100.00%        Mismatches:     0
Query Match:             100.00%        Indels:         0
DB:                      9              Gaps:           0

US-09-886-954A-1 (1-477) x BC019043 (1-2145)

Qy          1 MetThrProGluAspProGluGluThrGlnProLeuLeuGlyProProGlyGlySerAla 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         46 ATGACGCCCGAGGACCCAGAGGAAACCCAGCCGCTTCTGGGGCCTCCTGGCGGCAGCGCG 105

Qy         21 ProArgGlyArgArgValPheLeuAlaAlaPheAlaAlaAlaLeuGlyProLeuSerPhe 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        106 CCCCGCGGCCGCCGCGTCTTCCTCGCCGCCTTCGCCGCTGCCCTGGGCCCACTCAGCTTC 165

Qy         41 GlyPheAlaLeuGlyTyrSerSerProAlaIleProSerLeuGlnArgAlaAlaProPro 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        166 GGCTTCGCGCTCGGCTACAGCTCCCCGGCCATCCCTAGCCTGCAGCGCGCCGCGCCCCCG 225

Qy         61 AlaProArgLeuAspAspAlaAlaAlaSerTrpPheGlyAlaValValThrLeuGlyAla 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        226 GCCCCGCGCCTGGACGACGCCGCCGCCTCCTGGTTCGGGGCTGTCGTGACCCTGGGTGCC 285

Qy         81 AlaAlaGlyGlyValLeuGlyGlyTrpLeuValAspArgAlaGlyArgLysLeuSerLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        286 GCGGCGGGGGGAGTGCTGGGCGGCTGGCTGGTGGACCGCGCCGGGCGCAAGCTGAGCCTC 345

Qy        101 LeuLeuCysSerValProPheValAlaGlyPheAlaValIleThrAlaAlaGlnAspVal 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        346 TTGCTGTGCTCCGTGCCCTTCGTGGCCGGCTTTGCCGTCATCACCGCGGCCCAGGACGTG 405

Qy        121 TrpMetLeuLeuGlyGlyArgLeuLeuThrGlyLeuAlaCysGlyValAlaSerLeuVal 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        406 TGGATGCTGCTGGGGGGCCGCCTCCTCACCGGCCTGGCCTGCGGTGTTGCCTCCCTAGTG 465

Qy        141 AlaProValTyrIleSerGluIleAlaTyrProAlaValArgGlyLeuLeuGlySerCys 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        466 GCCCCGGTCTACATCTCCGAAATCGCCTACCCAGCAGTCCGGGGGTTGCTCGGCTCCTGT 525

Qy        161 ValGlnLeuMetValValValGlyIleLeuLeuAlaTyrLeuAlaGlyTrpValLeuGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        526 GTGCAGCTAATGGTCGTCGTCGGCATCCTCCTGGCCTACCTGGCAGGCTGGGTGCTGGAG 585

Qy        181 TrpArgTrpLeuAlaValLeuGlyCysValProProSerLeuMetLeuLeuLeuMetCys 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        586 TGGCGCTGGCTGGCTGTGCTGGGCTGCGTGCCCCCCTCCCTCATGCTGCTTCTCATGTGC 645

Qy        201 PheMetProGluThrProArgPheLeuLeuThrGlnHisArgArgGlnGluAlaMetAla 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        646 TTCATGCCCGAGACCCCGCGCTTCCTGCTGACTCAGCACAGGCGCCAGGAGGCCATGGCC 705
```

```
Qy         221 AlaLeuArgPheLeuTrpGlySerGluGlnGlyTrpGluAspProProIleGlyAlaGlu 240
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         706 GCCCTGCGGTTCCTGTGGGGCTCCGAGCAGGGCTGGGAAGACCCCCCCATCGGGGCTGAG 765

Qy         241 GlnSerPheHisLeuAlaLeuLeuArgGlnProGlyIleTyrLysProPheIleIleGly 260
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         766 CAGAGCTTTCACCTGGCCCTGCTGCGGCAGCCCGGCATCTACAAGCCCTTCATCATCGGC 825

Qy         261 ValSerLeuMetAlaPheGlnGlnLeuSerGlyValAsnAlaValMetPheTyrAlaGlu 280
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         826 GTCTCCCTGATGGCCTTCCAGCAGCTGTCGGGGGTCAACGCCGTCATGTTCTATGCAGAG 885

Qy         281 ThrIlePheGluGluAlaLysPheLysAspSerSerLeuAlaSerValValValGlyVal 300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         886 ACCATCTTTGAAGAGGCCAAGTTCAAGGACAGCAGCCTGGCCTCGGTCGTCGTGGGTGTC 945

Qy         301 IleGlnValLeuPheThrAlaValAlaAlaLeuIleMetAspArgAlaGlyArgArgLeu 320
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         946 ATCCAGGTGCTGTTCACAGCTGTGGCGGCTCTCATCATGGACAGAGCAGGGCGGAGGCTG 1005

Qy         321 LeuLeuValLeuSerGlyValValMetValPheSerThrSerAlaPheGlyAlaTyrPhe 340
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1006 CTCCTGGTCTTGTCAGGTGTGGTCATGGTGTTCAGCACGAGTGCCTTCGGCGCCTACTTC 1065

Qy         341 LysLeuThrGlnGlyGlyProGlyAsnSerSerHisValAlaIleSerAlaProValSer 360
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1066 AAGCTGACCCAGGGTGGCCCTGGCAACTCCTCGCACGTGGCCATCTCGGCGCCTGTCTCT 1125

Qy         361 AlaGlnProValAspAlaSerValGlyLeuAlaTrpLeuAlaValGlySerMetCysLeu 380
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1126 GCACAGCCTGTTGATGCCAGCGTGGGGCTGGCCTGGCTGGCCGTGGGCAGCATGTGCCTC 1185

Qy         381 PheIleAlaGlyPheAlaValGlyTrpGlyProIleProTrpLeuLeuMetSerGluIle 400
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1186 TTCATCGCCGGCTTTGCGGTGGGCTGGGGGCCCATCCCCTGGCTCCTCATGTCAGAGATC 1245

Qy         401 PheProLeuHisValLysGlyValAlaThrGlyIleCysValLeuThrAsnTrpLeuMet 420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1246 TTCCCTCTGCATGTCAAGGGCGTGGCGACAGGCATCTGCGTCCTCACCAACTGGCTCATG 1305

Qy         421 AlaPheLeuValThrLysGluPheSerSerLeuMetGluValLeuArgProTyrGlyAla 440
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1306 GCCTTTCTCGTGACCAAGGAGTTCAGCAGCCTCATGGAGGTCCTCAGGCCCTATGGAGCC 1365

Qy         441 PheTrpLeuAlaSerAlaPheCysIlePheSerValLeuPheThrLeuPheCysValPro 460
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1366 TTCTGGCTTGCCTCCGCTTTCTGCATCTTCAGTGTCCTTTTCACTTTGTTCTGTGTCCCT 1425

Qy         461 GluThrLysGlyLysThrLeuGluGlnIleThrAlaHisPheGluGlyArg 477
               |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1426 GAAACTAAAGGAAAGACTCTGGAACAAATCACAGCCCATTTTGAGGGGCGA 1476

RESULT 3
HSA245937
LOCUS       HSA245937                1873 bp    mRNA    linear   PRI 18-FEB-2000
DEFINITION  Homo sapiens mRNA for glucose transporter (GLUTX1 gene).
ACCESSION   AJ245937
VERSION     AJ245937.1  GI:7018305
KEYWORDS    glucose transporter; GLUTX1 gene.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Ibberson,M., Uldry,M. and Thorens,B.
  TITLE     GLUTX1, a novel mammalian glucose transporter expressed in the
            central nervous system and insulin-sensitive tissues
  JOURNAL   J. Biol. Chem. 275 (7), 4607-4612 (2000)
  MEDLINE   20138191
   PUBMED   10671487
REFERENCE   2  (bases 1 to 1873)
  AUTHORS   Ibberson,M.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-SEP-1999) Ibberson M.R., Institute of Pharmacology
            and Toxicology, University of Lausanne, Rue du Bugnon 27, Lausanne,
            CH-1005, SWITZERLAND
FEATURES             Location/Qualifiers
     source          1..1873
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
     gene            1..1873
                     /gene="GLUTX1"
     CDS             4..1437
                     /gene="GLUTX1"
                     /codon_start=1
                     /product="glucose transporter"
                     /protein_id="CAB75702.1"
                     /db_xref="GI:7018306"
                     /db_xref="GOA:Q9NY64"
                     /db_xref="SWISS-PROT:Q9NY64"
                     /translation="MTPEDPEETQPLLGPPGGSAPRGRRVFLAAFAAALGPLSFGFAL
                     GYSSPAIPSLQRAAPPAPRLDDAAASWFGAVVTLGAAAGGVLGGWLVDRAGRKLSLLL
                     CSVPFVAGFAVITAAQDVWMLLGGRLLTGLACGVASLVAPVYISEIAYPAVRGLLGSC
                     VQLMVVVGILLAYLAGWVLEWRWLAVLGCVPPSLMLLLMCFMPETPRFLLTQHRRQEA
                     MAALRFLWGSEQGWEDPPIGAEQSFHLALLRQPGIYKPFIIGVSLMAFQQLSGVNAVM
                     FYAETIFEEAKFKDSSLASVVVGVIQVLFTAVAALIMDRAGRRLLLVLSGVVMVFSTS
                     AFGAYFKLTQGGPGNSSHVAISAPVSAQPVDASVGLAWLAVGNMCLFIAGFAVGWGPI
                     PWLLMSEIFPLHVKGVATGICVLTNWLMAFLVTKEFSSLMEVLRPYGAFWLASAFCIF
                     SVLFTLFCVPEIKGKTLEQITAHFEGR"
ORIGIN

Alignment Scores:
Pred. No.:              9.55e-167      Length:          1873
Score:                  2448.00        Matches:         475
Percent Similarity:     99.79%         Conservative:    1
Best Local Similarity:  99.58%         Mismatches:      1
Query Match:            99.63%         Indels:          0
DB:                     9              Gaps:            0

US-09-886-954A-1 (1-477) x HSA245937 (1-1873)

Qy           1 MetThrProGluAspProGluGluThrGlnProLeuLeuGlyProProGlyGlySerAla 20
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           4 ATGACGCCCGAGGACCCAGAGGAAACCCAGCCGCTTCTGGGGCCTCCTGGCGGCAGCGCG 63

Qy          21 ProArgGlyArgArgValPheLeuAlaAlaPheAlaAlaAlaLeuGlyProLeuSerPhe 40
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          64 CCCCGCGGCCGCCGCGTCTTCCTCGCCGCCTTCGCCGCTGCCCTGGGCCCACTCAGCTTC 123

Qy          41 GlyPheAlaLeuGlyTyrSerSerProAlaIleProSerLeuGlnArgAlaAlaProPro 60
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         124 GGCTTCGCGCTCGGCTACAGCTCCCCGGCCATCCCTAGCCTGCAGCGCGCCGCGCCCCCG 183

Qy          61 AlaProArgLeuAspAspAlaAlaAlaSerTrpPheGlyAlaValValThrLeuGlyAla 80
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         184 GCCCCGCGCCTGGACGACGCCGCCGCCTCCTGGTTCGGGGCTGTCGTGACCCTGGGTGCC 243

Qy          81 AlaAlaGlyGlyValLeuGlyGlyTrpLeuValAspArgAlaGlyArgLysLeuSerLeu 100
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         244 GCGGCGGGGGGAGTGCTGGGCGGCTGGCTGGTGGACCGCGCCGGGCGCAAGCTGAGCCTC 303

Qy         101 LeuLeuCysSerValProPheValAlaGlyPheAlaValIleThrAlaAlaGlnAspVal 120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         304 TTGCTGTGCTCCGTGCCCTTCGTGGCCGGCTTTGCCGTCATCACCGCGGCCCAGGACGTG 363

Qy         121 TrpMetLeuLeuGlyGlyArgLeuLeuThrGlyLeuAlaCysGlyValAlaSerLeuVal 140
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         364 TGGATGCTGCTGGGGGGCCGCCTCCTCACCGGCCTGGCCTGCGGTGTTGCCTCCCTAGTG 423

Qy         141 AlaProValTyrIleSerGluIleAlaTyrProAlaValArgGlyLeuLeuGlySerCys 160
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         424 GCCCCGGTCTACATCTCCGAAATCGCCTACCCAGCAGTCCGGGGGTTGCTCGGCTCCTGT 483

Qy         161 ValGlnLeuMetValValValGlyIleLeuLeuAlaTyrLeuAlaGlyTrpValLeuGlu 180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db         484 GTGCAGCTAATGGTCGTCGTCGGCATCCTCCTGGCCTACCTGGCAGGCTGGGTGCTGGAG 543

Qy         181 TrpArgTrpLeuAlaValLeuGlyCysValProProSerLeuMetLeuLeuLeuMetCys 200
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         544 TGGCGCTGGCTGGCTGTGCTGGGCTGCGTGCCCCCCTCCCTCATGCTGCTTCTCATGTGC 603

Qy         201 PheMetProGluThrProArgPheLeuLeuThrGlnHisArgArgGlnGluAlaMetAla 220
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         604 TTCATGCCCGAGACCCCGCGCTTCCTGCTGACTCAGCACAGGCGCCAGGAGGCCATGGCC 663

Qy         221 AlaLeuArgPheLeuTrpGlySerGluGlnGlyTrpGluAspProProIleGlyAlaGlu 240
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         664 GCCCTGCGGTTCCTGTGGGGCTCCGAGCAGGGCTGGGAAGACCCCCCCATCGGGGCTGAG 723

Qy         241 GlnSerPheHisLeuAlaLeuLeuArgGlnProGlyIleTyrLysProPheIleIleGly 260
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         724 CAGAGCTTTCACCTGGCCCTGCTGCGGCAGCCCGGCATCTACAAGCCCTTCATCATCGGC 783

Qy         261 ValSerLeuMetAlaPheGlnGlnLeuSerGlyValAsnAlaValMetPheTyrAlaGlu 280
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         784 GTCTCCCTGATGGCCTTCCAGCAGCTGTCGGGGGTCAACGCCGTCATGTTCTATGCAGAG 843

Qy         281 ThrIlePheGluGluAlaLysPheLysAspSerSerLeuAlaSerValValValGlyVal 300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         844 ACCATCTTTGAAGAGGCCAAGTTCAAGGACAGCAGCCTGGCCTCGGTCGTCGTGGGTGTC 903

Qy         301 IleGlnValLeuPheThrAlaValAlaAlaLeuIleMetAspArgAlaGlyArgArgLeu 320
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         904 ATCCAGGTGCTGTTCACAGCTGTGGCGGCTCTCATCATGGACAGAGCAGGGCGGAGGCTG 963

Qy         321 LeuLeuValLeuSerGlyValValMetValPheSerThrSerAlaPheGlyAlaTyrPhe 340
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         964 CTCCTGGTCTTGTCAGGTGTGGTCATGGTGTTCAGCACGAGTGCCTTCGGCGCCTACTTC 1023

Qy         341 LysLeuThrGlnGlyGlyProGlyAsnSerSerHisValAlaIleSerAlaProValSer 360
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1024 AAGCTGACCCAGGGTGGCCCTGGCAACTCCTCGCACGTGGCCATCTCGGCGCCTGTCTCT 1083

Qy         361 AlaGlnProValAspAlaSerValGlyLeuAlaTrpLeuAlaValGlySerMetCysLeu 380
               ||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db        1084 GCACAGCCTGTTGATGCCAGCGTGGGGCTGGCCTGGCTGGCCGTGGGCAACATGTGCCTC 1143

Qy         381 PheIleAlaGlyPheAlaValGlyTrpGlyProIleProTrpLeuLeuMetSerGluIle 400
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1144 TTCATCGCCGGCTTTGCGGTGGGCTGGGGGCCCATCCCCTGGCTCCTCATGTCAGAGATC 1203

Qy         401 PheProLeuHisValLysGlyValAlaThrGlyIleCysValLeuThrAsnTrpLeuMet 420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1204 TTCCCTCTGCATGTCAAGGGCGTGGCGACAGGCATCTGCGTCCTCACCAACTGGCTCATG 1263

Qy         421 AlaPheLeuValThrLysGluPheSerSerLeuMetGluValLeuArgProTyrGlyAla 440
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1264 GCCTTTCTCGTGACCAAGGAGTTCAGCAGCCTCATGGAGGTCCTCAGGCCCTATGGAGCC 1323

Qy         441 PheTrpLeuAlaSerAlaPheCysIlePheSerValLeuPheThrLeuPheCysValPro 460
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1324 TTCTGGCTTGCCTCCGCTTTCTGCATCTTCAGTGTCCTTTTCACTTTGTTCTGTGTCCCT 1383

Qy         461 GluThrLysGlyLysThrLeuGluGlnIleThrAlaHisPheGluGlyArg 477
               |||   |||||||||||||||||||||||||||||||||||||||||||||
Db        1384 GAAATTAAAGGAAAGACTCTGGAACAAATCACAGCCCATTTTGAGGGGCGA 1434

RESULT 4
AX076667
LOCUS       AX076667                2217 bp    DNA     linear   PAT 06-FEB-2001
DEFINITION  Sequence 1 from Patent WO0104145.
ACCESSION   AX076667
VERSION     AX076667.1  GI:12711198
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Thorens,B., Ibberson,M. and Uldry,M.
  TITLE     Glutx polypeptide family and nucleic acids encoding same
  JOURNAL   Patent: WO 0104145-A 1 18-JAN-2001;
            University of Lausanne (CH)
FEATURES             Location/Qualifiers
     source          1. .2217
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
     CDS             348. .1781
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAC28495.1"
                     /db_xref="GI:12711199"
                     /db_xref="REMTREMBL:CAC28495"
                     /translation="MTPEDPEETQPLLGPPGGSAPRGRRVFLAAFAAALGPLSFGFAL
                     GYSSPAIPSLQRAAPPAPRLDDAAASWFGAVVTLGAAAGGVLGGWLVDRAGRKLSLLL
                     CSVPFVAGFAVITAAQDVWMLLGGRLLTGLACGVASLVAPVYISEIAYPAVRGLLGSC
                     VQLMVVVGILLAYLAGWVLEWRWLAVLGCVPPSLMLLLMCFMPETPRFLLTQHRRQEA
                     MAALRFLWGSEQGWEDPPIGAEQSFHLALLRQPGIYKPFIIGVSLMAFQQLSGVNAVM
                     FYAETIFEEAKFKDSSLASVVVGVIQVLFTAVAALIMDRAGRRLLLVLSGVVMVFSTS
                     AFGAYFKLTQGGPGNSSHVAISAPVSAQPVDASVGLAWLAVGNMCLFIAGFAVGWGPI
                     PWLLMSEIFPLHVKGVATGICVLTNWLMAFLVTKEFSSLMEVLRPYGAFWLASAFCIF
                     SVLFTLFCVPEIKGKTLEQITAHFEGR"
ORIGIN

Alignment Scores:
Pred. No.:              1.17e-166       Length:         2217
Score:                  2448.00         Matches:        475
Percent Similarity:     99.79%          Conservative:   1
Best Local Similarity:  99.58%          Mismatches:     1
Query Match:            99.63%          Indels:         0
DB:                     6               Gaps:           0

US-09-886-954A-1 (1-477) x AX076667 (1-2217)

Qy           1 MetThrProGluAspProGluGluThrGlnProLeuLeuGlyProProGlyGlySerAla 20
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         348 ATGACGCCCGAGGACCCAGAGGAAACCCAGCCGCTTCTGGGGCCTCCTGGCGGCAGCGCG 407

Qy          21 ProArgGlyArgArgValPheLeuAlaAlaPheAlaAlaAlaLeuGlyProLeuSerPhe 40
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         408 CCCCGCGGCCGCCGCGTCTTCCTCGCCGCCTTCGCCGCTGCCCTGGGCCCACTCAGCTTC 467

Qy          41 GlyPheAlaLeuGlyTyrSerSerProAlaIleProSerLeuGlnArgAlaAlaProPro 60
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         468 GGCTTCGCGCTCGGCTACAGCTCCCCGGCCATCCCTAGCCTGCAGCGCGCCGCGCCCCCG 527

Qy          61 AlaProArgLeuAspAspAlaAlaAlaSerTrpPheGlyAlaValValThrLeuGlyAla 80
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         528 GCCCCGCGCCTGGACGACGCCGCCGCCTCCTGGTTCGGGGCTGTCGTGACCCTGGGTGCC 587

Qy          81 AlaAlaGlyGlyValLeuGlyGlyTrpLeuValAspArgAlaGlyArgLysLeuSerLeu 100
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         588 GCGGCGGGGGGAGTGCTGGGCGGCTGGCTGGTGGACCGCGCCGGGCGCAAGCTGAGCCTC 647

Qy         101 LeuLeuCysSerValProPheValAlaGlyPheAlaValIleThrAlaAlaGlnAspVal 120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         648 TTGCTGTGCTCCGTGCCCTTCGTGGCCGGCTTTGCCGTCATCACCGCGGCCCAGGACGTG 707

Qy         121 TrpMetLeuLeuGlyGlyArgLeuLeuThrGlyLeuAlaCysGlyValAlaSerLeuVal 140
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         708 TGGATGCTGCTGGGGGGCCGCCTCCTCACCGGCCTGGCCTGCGGTGTTGCCTCCCTAGTG 767

Qy         141 AlaProValTyrIleSerGluIleAlaTyrProAlaValArgGlyLeuLeuGlySerCys 160
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         768 GCCCCGGTCTACATCTCCGAAATCGCCTACCCAGCAGTCCGGGGGTTGCTCGGCTCCTGT 827

Qy         161 ValGlnLeuMetValValValGlyIleLeuLeuAlaTyrLeuAlaGlyTrpValLeuGlu 180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         828 GTGCAGCTAATGGTCGTCGTCGGCATCCTCCTGGCCTACCTGGCAGGCTGGGTGCTGGAG 887
```

```
Qy        181 TrpArgTrpLeuAlaValLeuGlyCysValProProSerLeuMetLeuLeuLeuMetCys 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        888 TGGCGCTGGCTGGCTGTGCTGGGCTGCGTGCCCCCCTCCCTCATGCTGCTTCTCATGTGC 947

Qy        201 PheMetProGluThrProArgPheLeuLeuThrGlnHisArgArgGlnGluAlaMetAla 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        948 TTCATGCCCGAGACCCCGCGCTTCCTGCTGACTCAGCACAGGCGCCAGGAGGCCATGGCC 1007

Qy        221 AlaLeuArgPheLeuTrpGlySerGluGlnGlyTrpGluAspProProIleGlyAlaGlu 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1008 GCCCTGCGGTTCCTGTGGGGCTCCGAGCAGGGCTGGGAAGACCCCCCCATCGGGGCTGAG 1067

Qy        241 GlnSerPheHisLeuAlaLeuLeuArgGlnProGlyIleTyrLysProPheIleIleGly 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1068 CAGAGCTTTCACCTGGCCCTGCTGCGGCAGCCCGGCATCTACAAGCCCTTCATCATCGGC 1127

Qy        261 ValSerLeuMetAlaPheGlnGlnLeuSerGlyValAsnAlaValMetPheTyrAlaGlu 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1128 GTCTCCCTGATGGCCTTCCAGCAGCTGTCGGGGGTCAACGCCGTCATGTTCTATGCAGAG 1187

Qy        281 ThrIlePheGluGluAlaLysPheLysAspSerSerLeuAlaSerValValValGlyVal 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1188 ACCATCTTTGAAGAGGCCAAGTTCAAGGACAGCAGCCTGGCCTCGGTCGTCGTGGGTGTC 1247

Qy        301 IleGlnValLeuPheThrAlaValAlaAlaLeuIleMetAspArgAlaGlyArgArgLeu 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1248 ATCCAGGTGCTGTTCACAGCTGTGGCGGCTCTCATCATGGACAGAGCAGGGCGGAGGCTG 1307

Qy        321 LeuLeuValLeuSerGlyValValMetValPheSerThrSerAlaPheGlyAlaTyrPhe 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1308 CTCCTGGTCTTGTCAGGTGTGGTCATGGTGTTCAGCACGAGTGCCTTCGGCGCCTACTTC 1367

Qy        341 LysLeuThrGlnGlyGlyProGlyAsnSerSerHisValAlaIleSerAlaProValSer 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1368 AAGCTGACCCAGGGTGGCCCTGGCAACTCCTCGCACGTGGCCATCTCGGCGCCTGTCTCT 1427

Qy        361 AlaGlnProValAspAlaSerValGlyLeuAlaTrpLeuAlaValGlySerMetCysLeu 380
              ||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db       1428 GCACAGCCTGTTGATGCCAGCGTGGGGCTGGCCTGGCTGGCCGTGGGCAACATGTGCCTC 1487

Qy        381 PheIleAlaGlyPheAlaValGlyTrpGlyProIleProTrpLeuLeuMetSerGluIle 400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1488 TTCATCGCCGGCTTTGCGGTGGGCTGGGGGCCCATCCCCTGGCTCCTCATGTCAGAGATC 1547

Qy        401 PheProLeuHisValLysGlyValAlaThrGlyIleCysValLeuThrAsnTrpLeuMet 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1548 TTCCCTCTGCATGTCAAGGGCGTGGCGACAGGCATCTGCGTCCTCACCAACTGGCTCATG 1607

Qy        421 AlaPheLeuValThrLysGluPheSerSerLeuMetGluValLeuArgProTyrGlyAla 440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1608 GCCTTTCTCGTGACCAAGGAGTTCAGCAGCCTCATGGAGGTCCTCAGGCCCTATGGAGCC 1667

Qy        441 PheTrpLeuAlaSerAlaPheCysIlePheSerValLeuPheThrLeuPheCysValPro 460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1668 TTCTGGCTTGCCTCCGCTTTCTGCATCTTCAGTGTCCTTTTCACTTTGTTCTGTGTCCCT 1727

Qy        461 GluThrLysGlyLysThrLeuGluGlnIleThrAlaHisPheGluGlyArg 477
              |||   |||||||||||||||||||||||||||||||||||||||||||||
Db       1728 GAAATTAAAGGAAAGACTCTGGAACAAATCACAGCCCATTTTGAGGGGCGA 1778

RESULT 5
HSA17801
LOCUS       HSA17801                1508 bp    mRNA    linear   PRI 13-MAY-2000
DEFINITION  Homo sapiens mRNA for glucose transporter 8 (GLUT8 gene).
ACCESSION   Y17801
VERSION     Y17801.1  GI:7688145
KEYWORDS    glucose transporter 8; GLUT8 gene.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Doege,H., Schurmann,A., Bahrenberg,G., Brauers,A. and Joost,H.G.
  TITLE     GLUT8, a novel member of the sugar transport facilitator family
            with glucose transport activity
  JOURNAL   J. Biol. Chem. 275 (21), 16275-16280 (2000)
  MEDLINE   20283667
   PUBMED   10821868
REFERENCE   2  (bases 1 to 1508)
  AUTHORS   Joost,H.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JUL-1998) H. Joost, Institute of Pharmacology and
            Toxicology, Technical University Aachen, Wendlingweg 2, D-52057
            Aachen, FRG
FEATURES             Location/Qualifiers
     source          1. .1508
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /tissue_type="testis"
     gene            1. .1508
                     /gene="GLUT8"
     CDS             27. .1460
                     /gene="GLUT8"
                     /codon_start=1
                     /product="glucose transporter 8"
                     /protein_id="CAB89809.1"
                     /db_xref="GI:7688146"
                     /db_xref="GOA:Q9NY64"
                     /db_xref="SWISS-PROT:Q9NY64"
                     /translation="MTPEDPEETQPLLGPPGGSAPRGRRVFLAAFAAALGPLSFGFAL
                     GYSSPAIPSLQRAAPPAPRLDDAAASWFGAVVTLGAAAGGVLGGWLVDRAGRKLSLLL
                     CSVPFVAGFAVITAAQDVWMLLGGRLLTGLACGVASLVAPVYISEIAYPAVRGLLGSC
                     VQLMVVVGILLAYLAGWVLEWRWLAVLGCVPPSLMLLLMCFMPETPRFLLTQHRRQEA
                     MAALRFLWGSEQGWEDPPIGAEQSFHLALLRQPGIYKPFIIGVSLMAFQQLSGVNAVM
                     FYAETIFEEAKFKDSSLASVVVGVIQVLFTAVAALIMDRAGRRLLLVLSGVVMVFSTS
                     AFGAYFKLTQGGPGNSSHVAISAPVSAQPVDASVGLAWLAVGSMCLFIAGFAVGWGPI
                     PWLLMSEIFPLHVKGVATGICVLTNWLMAFLVTKEFSSLMEVLRPYGAFWLASAFCIF
                     SVLFTFSCVPETKGKTLEQITAHFEGR"
ORIGIN

Alignment Scores:
Pred. No.:              1.21e-166       Length:          1508
Score:                  2445.00         Matches:         475
Percent Similarity:     99.58%          Conservative:    0
Best Local Similarity:  99.58%          Mismatches:      2
Query Match:            99.51%          Indels:          0
DB:                     9               Gaps:            0

US-09-886-954A-1 (1-477) x HSA17801 (1-1508)

Qy          1 MetThrProGluAspProGluGluThrGlnProLeuLeuGlyProProGlyGlySerAla 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         27 ATGACGCCCGAGGACCCAGAGGAAACCCAGCCGCTTCTGGGGCCTCCTGGCGGCAGCGCG 86

Qy         21 ProArgGlyArgArgValPheLeuAlaAlaPheAlaAlaAlaLeuGlyProLeuSerPhe 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         87 CCCCGCGGCCGCCGCGTCTTCCTCGCCGCCTTCGCCGCTGCCCTGGGCCCACTCAGCTTC 146

Qy         41 GlyPheAlaLeuGlyTyrSerSerProAlaIleProSerLeuGlnArgAlaAlaProPro 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        147 GGCTTCGCGCTCGGCTACAGCTCCCCGGCCATCCCTAGCCTGCAGCGCGCCGCGCCCCCG 206

Qy         61 AlaProArgLeuAspAspAlaAlaAlaSerTrpPheGlyAlaValValThrLeuGlyAla 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        207 GCCCCGCGCCTGGACGACGCCGCCGCCTCCTGGTTCGGGGCTGTCGTGACCCTGGGTGCC 266

Qy         81 AlaAlaGlyGlyValLeuGlyGlyTrpLeuValAspArgAlaGlyArgLysLeuSerLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        267 GCGGCGGGGGGAGTGCTGGGCGGCTGGCTGGTGGACCGCGCCGGGCGCAAGCTGAGCCTC 326

Qy        101 LeuLeuCysSerValProPheValAlaGlyPheAlaValIleThrAlaAlaGlnAspVal 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        327 TTGCTGTGCTCCGTGCCCTTCGTGGCCGGCTTTGCCGTCATCACCGCGGCCCAGGACGTG 386

Qy        121 TrpMetLeuLeuGlyGlyArgLeuLeuThrGlyLeuAlaCysGlyValAlaSerLeuVal 140
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         387 TGGATGCTGCTGGGGGGCCGCCTCCTCACCGGCCTGGCCTGCGGTGTTGCCTCCCTAGTG 446

Qy         141 AlaProValTyrIleSerGluIleAlaTyrProAlaValArgGlyLeuLeuGlySerCys 160
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         447 GCCCCGGTCTACATCTCCGAAATCGCCTACCCAGCAGTCCGGGGGTTGCTCGGCTCCTGT 506

Qy         161 ValGlnLeuMetValValValGlyIleLeuLeuAlaTyrLeuAlaGlyTrpValLeuGlu 180
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         507 GTGCAGCTAATGGTCGTCGTCGGCATCCTCCTGGCCTACCTGGCAGGCTGGGTGCTGGAG 566

Qy         181 TrpArgTrpLeuAlaValLeuGlyCysValProProSerLeuMetLeuLeuLeuMetCys 200
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         567 TGGCGCTGGCTGGCTGTGCTGGGCTGCGTGCCCCCCTCCCTCATGCTGCTTCTCATGTGC 626

Qy         201 PheMetProGluThrProArgPheLeuLeuThrGlnHisArgArgGlnGluAlaMetAla 220
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         627 TTCATGCCCGAGACCCCGCGCTTCCTGCTGACTCAGCACAGGCGCCAGGAGGCCATGGCC 686

Qy         221 AlaLeuArgPheLeuTrpGlySerGluGlnGlyTrpGluAspProProIleGlyAlaGlu 240
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         687 GCCCTGCGGTTCCTGTGGGGCTCCGAGCAGGGCTGGAAGACCCCCCCATCGGGGCTGAG 746

Qy         241 GlnSerPheHisLeuAlaLeuLeuArgGlnProGlyIleTyrLysProPheIleIleGly 260
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         747 CAGAGCTTTCACCTGGCCCTGCTGCGGCAGCCCGGCATCTACAAGCCCTTCATCATCGGC 806

Qy         261 ValSerLeuMetAlaPheGlnGlnLeuSerGlyValAsnAlaValMetPheTyrAlaGlu 280
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         807 GTCTCCCTGATGGCCTTCCAGCAGCTGTCGGGGGTCAACGCCGTCATGTTCTATGCAGAG 866

Qy         281 ThrIlePheGluGluAlaLysPheLysAspSerSerLeuAlaSerValValValGlyVal 300
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         867 ACCATCTTTGAAGAGGCCAAGTTCAAGGACAGCAGCCTGGCCTCGGTCGTCGTGGGTGTC 926

Qy         301 IleGlnValLeuPheThrAlaValAlaAlaLeuIleMetAspArgAlaGlyArgArgLeu 320
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         927 ATCCAGGTGCTGTTCACAGCTGTGGCGGCTCTCATCATGGACAGAGCAGGGCGGAGGCTG 986

Qy         321 LeuLeuValLeuSerGlyValValMetValPheSerThrSerAlaPheGlyAlaTyrPhe 340
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         987 CTCCTGGTCTTGTCAGGTGTGGTCATGGTGTTCAGCACGAGTGCCTTCGGCGCCTACTTC 1046

Qy         341 LysLeuThrGlnGlyGlyProGlyAsnSerSerHisValAlaIleSerAlaProValSer 360
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1047 AAGCTGACCCAGGGTGGCCCTGGCAACTCCTCGCACGTGGCCATCTCGGCGCCTGTCTCT 1106

Qy         361 AlaGlnProValAspAlaSerValGlyLeuAlaTrpLeuAlaValGlySerMetCysLeu 380
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1107 GCACAGCCTGTTGATGCCAGCGTGGGGCTGGCCTGGCTGGCCGTGGGCAGCATGTGCCTC 1166

Qy         381 PheIleAlaGlyPheAlaValGlyTrpGlyProIleProTrpLeuLeuMetSerGluIle 400
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1167 TTCATCGCCGGCTTTGCGGTGGGCTGGGGGCCCATCCCCTGGCTCCTCATGTCAGAGATC 1226

Qy         401 PheProLeuHisValLysGlyValAlaThrGlyIleCysValLeuThrAsnTrpLeuMet 420
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1227 TTCCCTCTGCATGTCAAGGGCGTGGCGACAGGCATCTGCGTCCTCACCAACTGGCTCATG 1286

Qy         421 AlaPheLeuValThrLysGluPheSerSerLeuMetGluValLeuArgProTyrGlyAla 440
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1287 GCCTTTCTCGTGACCAAGGAGTTCAGCAGCCTCATGGAGGTCCTCAGGCCCTATGGAGCC 1346

Qy         441 PheTrpLeuAlaSerAlaPheCysIlePheSerValLeuPheThrLeuPheCysValPro 460
                |||||||||||||||||||||||||||||||||||||||||||||      |||||||||
Db        1347 TTCTGGCTTGCCTCCGCTTTCTGCATCTTCAGTGTCCTTTTCACTTTTTCCTGTGTCCCT 1406

Qy         461 GluThrLysGlyLysThrLeuGluGlnIleThrAlaHisPheGluGlyArg 477
                |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1407 GAAACTAAAGGAAAGACTCTGGAACAAATCACAGCCCATTTTGAGGGGCGA 1457

RESULT 6
AX076669
LOCUS       AX076669                2087 bp    DNA     linear   PAT 06-FEB-2001
DEFINITION  Sequence 3 from Patent WO0104145.
ACCESSION   AX076669
VERSION     AX076669.1  GI:12711200
KEYWORDS    .
SOURCE      Rattus sp.
  ORGANISM  Rattus sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1
  AUTHORS   Thorens,B., Ibberson,M. and Uldry,M.
  TITLE     Glutx polypeptide family and nucleic acids encoding same
  JOURNAL   Patent: WO 0104145-A 3 18-JAN-2001;
            University of Lausanne (CH)
FEATURES             Location/Qualifiers
     source          1. .2087
                     /organism="Rattus sp."
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:10118"
     CDS             30. .1466
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAC28496.1"
                     /db_xref="GI:12711201"
                     /db_xref="REMTREMBL:CAC28496"
                     /translation="MSPEDPQETQPLLRSPGARAPGGRRVFLATFAAALGPLSFGFAL
                     GYSSPAIPSLRRTAPPALRLGDTAASWFGAVVTLGAAAGGVLGGWLLDRAGRKLSLLL
                     CTVPFVTGFAVITAARDVWMLLGGRLLTGLACGVASLVAPVYISEIAYPAVRGLLGSC
                     VQLMVVTGILLAYVAGWVLEWRWLAVLGCVPPTLMLLLMCYMPETPRFLLTQHQYQEA
                     MAALRFLWGSEEGWEEPPVGAEHQGFQLAMLRRPGVHKPLIIGICLMVFQQLSGVNAI
                     MFYANTIFEEAKFKDSSLASVTVGIIQVLFTAVAALIMDRAGRKLLLALSGVIMVFSM
                     SAFGTYFKLTQSGPSNSSHVGLLVPISAEPADVHLGLAWLAVGSMCLFIAGFAVGWGP
                     IPWLLMSEIFPLHIKGVATGVCVLTNWFMAFLVTKEFNSIMEILRPYGAFWLTAAFCI
                     LSVLFTLTFVPETKGRTLEQITAHFEGR"
ORIGIN

Alignment Scores:
Pred. No.:              3.51e-146      Length:         2087
Score:                  2162.50        Matches:        408
Percent Similarity:     92.26%         Conservative:   33
Best Local Similarity:  85.36%         Mismatches:     36
Query Match:            88.01%         Indels:         1
DB:                     6              Gaps:           1

US-09-886-954A-1 (1-477) x AX076669 (1-2087)

Qy           1 MetThrProGluAspProGluGluThrGlnProLeuLeuGlyProProGlyGlySerAla 20
                |||:::||||||||||||:::||||||||||||||||||      ||||||      |||
Db          30 ATGTCGCCCGAGGACCCCCAGGAAACACAGCCGCTGTTGCGGTCGCCGGGCGCCAGGGCT 89

Qy          21 ProArgGlyArgArgValPheLeuAlaAlaPheAlaAlaAlaLeuGlyProLeuSerPhe 40
                |||   |||||||||||||||||||||   ||||||||||||||||||||||||||||||
Db          90 CCCGGCGGCCGCCGGGTCTTCCTTGCCACCTTCGCCGCCGCCCTGGGCCCGCTCAGCTTC 149

Qy          41 GlyPheAlaLeuGlyTyrSerSerProAlaIleProSerLeuGlnArgAlaAlaProPro 60
                ||||||||||||||||||||||||||||||||||||||||||:::|||   |||||||||
Db         150 GGCTTCGCGCTTGGCTACAGCTCCCCCGCCATCCCCAGCCTGCGGCGCACCGCACCCCCT 209

Qy          61 AlaProArgLeuAspAspAlaAlaAlaSerTrpPheGlyAlaValValThrLeuGlyAla 80
                |||   ||||||   |||   |||||||||||||||||||||||||||||||||||||||
Db         210 GCCCTACGCCTCGGAGACACTGCGGCCTCCTGGTTCGGGGCCGTCGTGACCCTGGGCGCT 269

Qy          81 AlaAlaGlyGlyValLeuGlyGlyTrpLeuValAspArgAlaGlyArgLysLeuSerLeu 100
                ||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||
Db         270 GCGGCAGGGGGCGTGCTGGGCGGCTGGCTCCTGGACCGTGCAGGGCGCAAGCTGAGCCTC 329

Qy         101 LeuLeuCysSerValProPheValAlaGlyPheAlaValIleThrAlaAlaGlnAspVal 120
                |||||||||:::||||||||||||   ||||||||||||||||||||||||:::||||||
Db         330 CTGCTCTGCACCGTGCCCTTCGTGACCGGTTTTGCTGTCATCACCGCGGCCCGGGATGTG 389

Qy         121 TrpMetLeuLeuGlyGlyArgLeuLeuThrGlyLeuAlaCysGlyValAlaSerLeuVal 140
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        390 TGGATGCTGCTCGGAGGCCGCCTCCTCACCGGCCTAGCCTGCGGAGTCGCCTCATTAGTG 449

Qy        141 AlaProValTyrIleSerGluIleAlaTyrProAlaValArgGlyLeuLeuGlySerCys 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        450 GCACCGGTCTATATCTCGGAAATCGCCTACCCAGCCGTTCGAGGACTGCTCGGCTCCTGT 509

Qy        161 ValGlnLeuMetValValValGlyIleLeuLeuAlaTyrLeuAlaGlyTrpValLeuGlu 180
              ||||||||||||||||||   ||||||||||||||||||:::||||||||||||||||||
Db        510 GTGCAGCTGATGGTTGTCACTGGCATCCTCTTGGCCTACGTGGCAGGCTGGGTCCTAGAA 569

Qy        181 TrpArgTrpLeuAlaValLeuGlyCysValProProSerLeuMetLeuLeuLeuMetCys 200
              ||||||||||||||||||||||||||||||||||||:::|||||||||||||||||||||
Db        570 TGGCGCTGGTTGGCTGTGCTGGGCTGTGTGCCCCCCACCCTCATGCTGCTGCTCATGTGC 629

Qy        201 PheMetProGluThrProArgPheLeuLeuThrGlnHisArgArgGlnGluAlaMetAla 220
              :::||||||||||||||||||||||||||||||||||||:::   |||||||||||||||
Db        630 TACATGCCCGAGACCCCACGCTTCCTCCTGACGCAACACCAGTACCAGGAGGCCATGGCT 689

Qy        221 AlaLeuArgPheLeuTrpGlySerGluGlnGlyTrpGluAspProProIleGlyAlaGlu 240
              |||||||||||||||||||||||||||:::||||||||||||:::||||||:::|||||||||
Db        690 GCCCTGCGCTTCCTGTGGGGCTCTGAGGAGGGCTGGGAAGAGCCCCCTGTTGGGGCTGAG 749

Qy        241 ---GlnSerPheHisLeuAlaLeuLeuArgGlnProGlyIleTyrLysProPheIleIle 259
                 |||   |||   ||||||:::||||||:::||||||::::::||||||   ||||||
Db        750 CACCAGGGCTTCCAGCTGGCTATGCTGAGGCGCCCTGGTGTCCACAAGCCCCTCATCATC 809

Qy        260 GlyValSerLeuMetAlaPheGlnGlnLeuSerGlyValAsnAlaValMetPheTyrAla 279
              |||:::   ||||||   |||||||||||||||||||||||||||:::||||||||||||
Db        810 GGCATTTGCCTCATGGTCTTCCAGCAGCTGTCAGGGGTCAACGCCATCATGTTCTATGCC 869

Qy        280 GluThrIlePheGluGluAlaLysPheLysAspSerSerLeuAlaSerValValValGly 299
                 ||||||||||||||||||||||||||||||||||||||||||||||||   ||||||
Db        870 AACACCATCTTTGAGGAGGCCAAGTTCAAGGACAGCAGCCTGGCCTCGGTCACTGTGGGC 929

Qy        300 ValIleGlnValLeuPheThrAlaValAlaAlaLeuIleMetAspArgAlaGlyArgArg 319
              :::||||||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db        930 ATCATCCAGGTTCTGTTCACTGCTGTGGCGGCCCTCATCATGGACAGAGCAGGGCGAAAA 989

Qy        320 LeuLeuLeuValLeuSerGlyValValMetValPheSerThrSerAlaPheGlyAlaTyr 339
              |||||||||   ||||||||||||:::||||||||||||   ||||||||||||   |||
Db        990 CTGCTTCTGGCCTTGTCGGGTGTGATCATGGTTTTCAGCATGAGCGCCTTTGGTACCTAC 1049

Qy        340 PheLysLeuThrGlnGlyGlyProGlyAsnSerSerHisValAlaIleSerAlaProVal 359
              |||||||||||||||   ||||||   |||||||||||||||   :::      |||:::
Db       1050 TTCAAACTGACCCAGAGTGGCCCCAGCAACTCCTCCCATGTAGGCCTCCTGGTGCCCATC 1109

Qy        360 SerAlaGlnProValAspAlaSerValGlyLeuAlaTrpLeuAlaValGlySerMetCys 379
              ||||||:::|||   |||      :::|||||||||||||||||||||||||||||||||
Db       1110 TCCGCAGAGCCTGCTGATGTTCACCTGGGGCTGGCCTGGCTGGCTGTAGGCAGCATGTGC 1169

Qy        380 LeuPheIleAlaGlyPheAlaValGlyTrpGlyProIleProTrpLeuLeuMetSerGlu 399
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1170 CTCTTCATCGCTGGTTTTGCAGTAGGCTGGGGACCCATCCCCTGGCTCCTCATGTCAGAG 1229

Qy        400 IlePheProLeuHisValLysGlyValAlaThrGlyIleCysValLeuThrAsnTrpLeu 419
              |||||||||||||||:::||||||||||||||||||:::||||||||||||||||||
Db       1230 ATCTTCCCTCTGCACATCAAGGGTGTGGCTACCGGCGTCTGTGTCCTCACCAACTGGTTC 1289

Qy        420 MetAlaPheLeuValThrLysGluPheSerSerLeuMetGluValLeuArgProTyrGly 439
              |||||||||||||||||||||||||||:::|||:::||||||:::|||||||||||||||
Db       1290 ATGGCCTTTCTGGTGACCAAAGAGTTTAACAGCATCATGGAGATCCTCAGACCCTACGGC 1349

Qy        440 AlaPheTrpLeuAlaSerAlaPheCysIlePheSerValLeuPheThrLeuPheCysVal 459
              ||||||||||||   :::||||||||||||   ||||||||||||||||||      |||
Db       1350 GCCTTCTGGCTCACCGCTGCCTTCTGTATCCTCAGCGTCCTTTTCACGCTCACCTTTGTC 1409

Qy        460 ProGluThrLysGlyLysThrLeuGluGlnIleThrAlaHisPheGluGlyArg 477
              |||||||||||||||:::||||||||||||||||||||||||||||||||||||
Db       1410 CCTGAGACTAAAGGCAGGACTCTGGAACAAATCACAGCCCATTTCGAGGGACGG 1463

RESULT 7
RNO245935
LOCUS       RNO245935                2087 bp    mRNA    linear   ROD 18-FEB-2000
DEFINITION  Rattus norvegicus mRNA for glucose transporter (GLUTX1 gene).
ACCESSION   AJ245935
VERSION     AJ245935.1  GI:7018604
KEYWORDS    glucose transporter; GLUTX1 gene.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1
  AUTHORS   Ibberson,M., Uldry,M. and Thorens,B.
  TITLE     GLUTX1, a novel mammalian glucose transporter expressed in the
            central nervous system and insulin-sensitive tissues
  JOURNAL   J. Biol. Chem. 275 (7), 4607-4612 (2000)
  MEDLINE   20138191
   PUBMED   10671487
REFERENCE   2  (bases 1 to 2087)
  AUTHORS   Ibberson,M.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-SEP-1999) Ibberson M.R., Institute of Pharmacology
            and Toxicology, University of Lausanne, Rue du Bugnon 27, Lausanne,
            CH-1005, SWITZERLAND
FEATURES             Location/Qualifiers
     source          1..2087
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10116"
     gene            1..2087
                     /gene="GLUTX1"
     CDS             30..1466
                     /gene="GLUTX1"
                     /codon_start=1
                     /product="glucose transporter"
                     /protein_id="CAB75729.1"
                     /db_xref="GI:7018605"
                     /db_xref="GOA:Q9JJZ1"
                     /db_xref="SWISS-PROT:Q9JJZ1"
                     /translation="MSPEDPQETQPLLRSPGARAPGGRRVFLATFAAALGPLSFGFAL
                     GYSSPAIPSLRRTAPPALRLGDTAASWFGAVVTLGAAAGGVLGGWLLDRAGRKLSLLL
                     CTVPFVTGFAVITAARDVWMLLGGRLLTGLACGVASLVAPVYISEIAYPAVRGLLGSC
                     VQLMVVTGILLAYVAGWVLEWRWLAVLGCVPPTLMLLLMCYMPETPRFLLTQHQYQEA
                     MAALRFLWGSEEGWEEPPVGAEHQGFQLAMLRRPGVHKPLIIGICLMVFQQLSGVNAI
                     MFYANTIFEEAKFKDSSLASVTVGIIQVLFTAVAALIMDRAGRKLLLALSGVIMVFSM
                     SAFGTYFKLTQSGPSNSSHVGLLVPISAEPADVHLGLAWLAVGSMCLFIAGFAVGWGP
                     IPWLLMSEIFPLHIKGVATGVCVLTNWFMAFLVTKEFNSIMEILRPYGAFWLTAAFCI
                     LSVLFTLTFVPETKGRTLEQITAHFEGR"
ORIGIN

Alignment Scores:
Pred. No.:              3.51e-146      Length:          2087
Score:                  2162.50        Matches:         408
Percent Similarity:     92.26%         Conservative:    33
Best Local Similarity:  85.36%         Mismatches:      36
Query Match:            88.01%         Indels:          1
DB:                     10             Gaps:            1

US-09-886-954A-1 (1-477) x RNO245935 (1-2087)

Qy          1 MetThrProGluAspProGluGluThrGlnProLeuLeuGlyProProGlyGlySerAla 20
              |||:::||||||||||||:::||||||||||||||||||      ||||||      |||
Db         30 ATGTCGCCCGAGGACCCCCAGGAAACACAGCCGCTGTTGCGGTCGCCGGGCGCCAGGGCT 89

Qy         21 ProArgGlyArgArgValPheLeuAlaAlaPheAlaAlaAlaLeuGlyProLeuSerPhe 40
              |||   |||||||||||||||||||||   ||||||||||||||||||||||||||||||
Db         90 CCCGGCGGCCGCCGGGTCTTCCTTGCCACCTTCGCCGCCGCCCTGGGCCCGCTCAGCTTC 149

Qy         41 GlyPheAlaLeuGlyTyrSerSerProAlaIleProSerLeuGlnArgAlaAlaProPro 60
              ||||||||||||||||||||||||||||||||||||||||||:::|||   |||||||||
Db        150 GGCTTCGCGCTTGGCTACAGCTCCCCCGCCATCCCCAGCCTGCGGCGCACCGCACCCCCT 209

Qy         61 AlaProArgLeuAspAspAlaAlaAlaSerTrpPheGlyAlaValValThrLeuGlyAla 80
              |||   ||||||   |||   |||||||||||||||||||||||||||||||||||||||
Db        210 GCCCTACGCCTCGGAGACACTGCGGCCTCCTGGTTCGGGGCCGTCGTGACCCTGGGCGCT 269
```

```
Qy         81 AlaAlaGlyGlyValLeuGlyGlyTrpLeuValAspArgAlaGlyArgLysLeuSerLeu 100
              ||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||
Db        270 GCGGCAGGGGGCGTGCTGGGCGGCTGGCTCCTGGACCGTGCAGGGCGCAAGCTGAGCCTC 329

Qy        101 LeuLeuCysSerValProPheValAlaGlyPheAlaValIleThrAlaAlaGlnAspVal 120
              |||||||||:::||||||||||||   ||||||||||||||||||||||||:::||||||
Db        330 CTGCTCTGCACCGTGCCCTTCGTGACCGGTTTTGCTGTCATCACCGCGGCCCGGGATGTG 389

Qy        121 TrpMetLeuLeuGlyGlyArgLeuLeuThrGlyLeuAlaCysGlyValAlaSerLeuVal 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        390 TGGATGCTGCTCGGAGGCCGCCTCCTCACCGGCCTAGCCTGCGGAGTCGCCTCATTAGTG 449

Qy        141 AlaProValTyrIleSerGluIleAlaTyrProAlaValArgGlyLeuLeuGlySerCys 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        450 GCACCGGTCTATATCTCGGAAATCGCCTACCCAGCCGTTCGAGGACTGCTCGGCTCCTGT 509

Qy        161 ValGlnLeuMetValValValGlyIleLeuLeuAlaTyrLeuAlaGlyTrpValLeuGlu 180
              ||||||||||||||||||   ||||||||||||||||||:::||||||||||||||||||
Db        510 GTGCAGCTGATGGTTGTCACTGGCATCCTCTTGGCCTACGTGGCAGGCTGGGTCCTAGAA 569

Qy        181 TrpArgTrpLeuAlaValLeuGlyCysValProProSerLeuMetLeuLeuLeuMetCys 200
              ||||||||||||||||||||||||||||||||||||:::|||||||||||||||||||||
Db        570 TGGCGCTGGTTGGCTGTGCTGGGCTGTGTGCCCCCCACCCTCATGCTGCTGCTCATGTGC 629

Qy        201 PheMetProGluThrProArgPheLeuLeuThrGlnHisArgArgGlnGluAlaMetAla 220
              :::||||||||||||||||||||||||||||||||||||:::   |||||||||||||||
Db        630 TACATGCCCGAGACCCCACGCTTCCTCCTGACGCAACACCAGTACCAGGAGGCCATGGCT 689

Qy        221 AlaLeuArgPheLeuTrpGlySerGluGlnGlyTrpGluAspProProIleGlyAlaGlu 240
              |||||||||||||||||||||||||||:::|||||||||:::||||||:::|||||||||
Db        690 GCCCTGCGCTTCCTGTGGGGCTCTGAGGAGGGCTGGGAAGAGCCCCCTGTTGGGGCTGAG 749

Qy        241 ---GlnSerPheHisLeuAlaLeuLeuArgGlnProGlyIleTyrLysProPheIleIle 259
                 |||   |||   ||||||:::||||||:::||||||::::::||||||   ||||||
Db        750 CACCAGGGCTTCCAGCTGGCTATGCTGAGGCGCCCTGGTGTCCACAAGCCCCTCATCATC 809

Qy        260 GlyValSerLeuMetAlaPheGlnGlnLeuSerGlyValAsnAlaValMetPheTyrAla 279
              |||:::   ||||||   |||||||||||||||||||||||||||:::||||||||||||
Db        810 GGCATTTGCCTCATGGTCTTCCAGCAGCTGTCAGGGGTCAACGCCATCATGTTCTATGCC 869

Qy        280 GluThrIlePheGluGluAlaLysPheLysAspSerSerLeuAlaSerValValValGly 299
                 ||||||||||||||||||||||||||||||||||||||||||||||||   ||||||
Db        870 AACACCATCTTTGAGGAGGCCAAGTTCAAGGACAGCAGCCTGGCCTCGGTCACTGTGGGC 929

Qy        300 ValIleGlnValLeuPheThrAlaValAlaAlaLeuIleMetAspArgAlaGlyArgArg 319
              :::||||||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db        930 ATCATCCAGGTTCTGTTCACTGCTGTGGCGGCCCTCATCATGGACAGAGCAGGGCGAAAA 989

Qy        320 LeuLeuLeuValLeuSerGlyValValMetValPheSerThrSerAlaPheGlyAlaTyr 339
              |||||||||   ||||||||||||:::||||||||||||   ||||||||||||   |||
Db        990 CTGCTTCTGGCCTTGTCGGGTGTGATCATGGTTTTCAGCATGAGCGCCTTTGGTACCTAC 1049

Qy        340 PheLysLeuThrGlnGlyGlyProGlyAsnSerSerHisValAlaIleSerAlaProVal 359
              |||||||||||||||   ||||||   |||||||||||||||   :::      |||:::
Db       1050 TTCAAACTGACCCAGAGTGGCCCCAGCAACTCCTCCCATGTAGGCCTCCTGGTGCCCATC 1109

Qy        360 SerAlaGlnProValAspAlaSerValGlyLeuAlaTrpLeuAlaValGlySerMetCys 379
              ||||||:::|||   |||      :::|||||||||||||||||||||||||||||||||
Db       1110 TCCGCAGAGCCTGCTGATGTTCACCTGGGGCTGGCCTGGCTGGCTGTAGGCAGCATGTGC 1169

Qy        380 LeuPheIleAlaGlyPheAlaValGlyTrpGlyProIleProTrpLeuLeuMetSerGlu 399
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1170 CTCTTCATCGCTGGTTTTGCAGTAGGCTGGGGACCCATCCCCTGGCTCCTCATGTCAGAG 1229

Qy        400 IlePheProLeuHisValLysGlyValAlaThrGlyIleCysValLeuThrAsnTrpLeu 419
              |||||||||||||||:::||||||||||||||||||:::||||||||||||||||||
Db       1230 ATCTTCCCTCTGCACATCAAGGGTGTGGCTACCGGCGTCTGTGTCCTCACCAACTGGTTC 1289

Qy        420 MetAlaPheLeuValThrLysGluPheSerSerLeuMetGluValLeuArgProTyrGly 439
              |||||||||||||||||||||||||||:::|||:::||||||:::|||||||||||||||
Db       1290 ATGGCCTTTCTGGTGACCAAAGAGTTTAACAGCATCATGGAGATCCTCAGACCCTACGGC 1349
```

```
Qy        440 AlaPheTrpLeuAlaSerAlaPheCysIlePheSerValLeuPheThrLeuPheCysVal 459
              ||||||||||||   :::||||||||||||   ||||||||||||||||||      |||
Db       1350 GCCTTCTGGCTCACCGCTGCCTTCTGTATCCTCAGCGTCCTTTTCACGCTCACCTTTGTC 1409

Qy        460 ProGluThrLysGlyLysThrLeuGluGlnIleThrAlaHisPheGluGlyArg 477
              |||||||||||||||:::||||||||||||||||||||||||||||||||||||
Db       1410 CCTGAGACTAAAGGCAGGACTCTGGAACAAATCACAGCCCATTTCGAGGGACGG 1463

RESULT 8
MMU17802
LOCUS       MMU17802                1490 bp    mRNA    linear   ROD 13-MAY-2000
DEFINITION  Mus musculus mRNA for glucose transporter 8 (GLUT8 gene).
ACCESSION   Y17802
VERSION     Y17802.1  GI:7688219
KEYWORDS    glucose transporter 8; GLUT8 gene.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Doege,H., Schurmann,A., Bahrenberg,G., Brauers,A. and Joost,H.G.
  TITLE     GLUT8, a novel member of the sugar transport facilitator family
            with glucose transport activity
  JOURNAL   J. Biol. Chem. 275 (21), 16275-16280 (2000)
  MEDLINE   20283667
   PUBMED   10821868
REFERENCE   2  (bases 1 to 1490)
  AUTHORS   Joost,H.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JUL-1998) H. Joost, Institute of Pharmacology and
            Toxicology, Technical University Aachen, Wendlingweg 2, D-52057
            Aachen, FRG
FEATURES             Location/Qualifiers
     source          1..1490
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /tissue_type="testis"
     gene            1..1490
                     /gene="GLUT8"
     CDS             26..1459
                     /gene="GLUT8"
                     /codon_start=1
                     /product="glucose transporter 8"
                     /protein_id="CAB89815.1"
                     /db_xref="GI:7688220"
                     /db_xref="GOA:Q9JIF3"
                     /db_xref="SWISS-PROT:Q9JIF3"
                     /translation="MSPEDPQETQPLLRPPEARTPRGRRVFLASFAAALGPLSFGFAL
                     GYSSPAIPSLRRTAPPALRLGDNAASWFGAVVTLGAAAGGILGGWLLDRAGRKLSLLL
                     CTVPFVTGFAVITAARDVWMLLGGRLLTGLACGVASLVAPVYISEIAYPAVRGLLGSC
                     VQLMVVTGILLAYVAGWVLEWRWLAVLGCVPPTLMLLLMCYMPETPRFLLTQHQYQEA
                     MAALRFLWGSEEGWEEPPVGAEHQGFQLALLRRPGIYKPLIIGISLMVFQQLSGVNAI
                     MFYANSIFEEAKFKDSSLASVTVGIIQVLFTAVAALIMDRAGRRLLLALSGVIMVFSM
                     SAFGTYFKLTQSLPSNSSHVGLVPIAAEPVDVQVGLAWLAVGSMCLFIAGFAVGWGPI
                     PWLLMSEIFPLHVKGVATGICVLTNWFMAFLVTKEFSSVMEMLRPYGAFWLTAAFCAL
                     SVLFTLTVVPETKGRTLEQVTAHFEGR"
ORIGIN

Alignment Scores:
Pred. No.:              4.19e-146       Length:         1490
Score:                  2159.00         Matches:        412
Percent Similarity:     92.47%          Conservative:   30
Best Local Similarity:  86.19%          Mismatches:     34
Query Match:            87.87%          Indels:         2
DB:                     10              Gaps:           2

US-09-886-954A-1 (1-477) x MMU17802 (1-1490)

Qy          1 MetThrProGluAspProGluGluThrGlnProLeuLeuGlyProProGlyGlySerAla 20
              |||:::||||||||||||:::||||||||||||||||||   ||||||
Db         26 ATGTCTCCCGAGGACCCCCAGGAGACGCAGCCGCTATTGCGGCCACCGGAAGCCAGGACT 85
```

```
Qy       21 ProArgGlyArgArgValPheLeuAlaAlaPheAlaAlaAlaLeuGlyProLeuSerPhe 40
            |||||||||||||||||||||||||||:::||||||||||||||||||||||||||||||
Db       86 CCCCGCGGCCGCCGGGTCTTCCTCGCTTCCTTCGCCGCCGCTCTGGGACCCCTCAGCTTC 145

Qy       41 GlyPheAlaLeuGlyTyrSerSerProAlaIleProSerLeuGlnArgAlaAlaProPro 60
            ||||||||||||||||||||||||||||||||||||||||||:::|||   |||||||||
Db      146 GGCTTCGCGCTCGGCTACAGCTCCCCCGCCATCCCCAGCCTGCGGCGCACCGCACCCCCG 205

Qy       61 AlaProArgLeuAspAspAlaAlaAlaSerTrpPheGlyAlaValValThrLeuGlyAla 80
            |||   ||||||   |||   |||||||||||||||||||||||||||||||||||||||
Db      206 GCCCTGCGCCTCGGAGACAATGCGGCCTCCTGGTTCGGGGCCGTCGTGACCCTGGGCGCT 265

Qy       81 AlaAlaGlyGlyValLeuGlyGlyTrpLeuValAspArgAlaGlyArgLysLeuSerLeu 100
            ||||||||||||:::|||||||||||||||:::|||||||||||||||||||||||||||
Db      266 GCTGCAGGGGGCATACTGGGCGGCTGGCTCCTGGACCGTGCAGGGCGCAAGCTGAGCCTC 325

Qy      101 LeuLeuCysSerValProPheValAlaGlyPheAlaValIleThrAlaAlaGlnAspVal 120
            |||||||||:::||||||||||||   ||||||||||||||||||||||||:::||||||
Db      326 TTGCTCTGCACCGTGCCCTTCGTGACTGGCTTTGCTGTCATCACCGCGGCCCGGGATGTG 385

Qy      121 TrpMetLeuLeuGlyGlyArgLeuLeuThrGlyLeuAlaCysGlyValAlaSerLeuVal 140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      386 TGGATGCTGCTCGGAGGCCGCCTCCTCACCGGCCTTGCCTGCGGAGTCGCCTCACTAGTG 445

Qy      141 AlaProValTyrIleSerGluIleAlaTyrProAlaValArgGlyLeuLeuGlySerCys 160
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      446 GCACCGGTTTACATCTCGGAAATCGCCTACCCAGCTGTCCGAGGACTGCTCGGCTCCTGT 505

Qy      161 ValGlnLeuMetValValValGlyIleLeuLeuAlaTyrLeuAlaGlyTrpValLeuGlu 180
            ||||||||||||||||||   ||||||||||||||||||:::||||||||||||||||||
Db      506 GTGCAGCTGATGGTTGTCACTGGCATCCTCCTGGCCTATGTGGCAGGCTGGGTCCTAGAG 565

Qy      181 TrpArgTrpLeuAlaValLeuGlyCysValProProSerLeuMetLeuLeuLeuMetCys 200
            ||||||||||||||||||||||||||||||||||||:::|||||||||||||||||||||
Db      566 TGGCGCTGGCTGGCCGTGCTGGGCTGTGTGCCCCCCACCCTCATGCTGCTGCTCATGTGC 625

Qy      201 PheMetProGluThrProArgPheLeuLeuThrGlnHisArgArgGlnGluAlaMetAla 220
            :::||||||||||||||||||||||||||||||||||||:::   |||||||||||||||
Db      626 TACATGCCCGAAACCCCACGTTTTCTCCTCACTCAACACCAGTACCAGGAGGCCATGGCT 685

Qy      221 AlaLeuArgPheLeuTrpGlySerGluGlnGlyTrpGluAspProProIleGlyAlaGlu 240
            |||||||||||||||||||||||||||:::|||||||||:::||||||:::|||||||||
Db      686 GCCTTGCGCTTCCTGTGGGGCTCTGAGGAGGGCTGGGAAGAGCCCCCTGTTGGGGCTGAG 745

Qy      241 ---GlnSerPheHisLeuAlaLeuLeuArgGlnProGlyIleTyrLysProPheIleIle 259
               |||   |||   |||||||||||||||:::||||||||||||||||||   ||||||
Db      746 CACCAGGGCTTCCAGCTGGCCCTGCTGAGGCGCCCTGGCATCTACAAGCCCCTCATCATC 805

Qy      260 GlyValSerLeuMetAlaPheGlnGlnLeuSerGlyValAsnAlaValMetPheTyrAla 279
            |||:::|||||||||   |||||||||||||||||||||||||||:::||||||||||||
Db      806 GGCATTTCCCTCATGGTCTTCCAGCAGCTGTCAGGGGTCAATGCTATCATGTTCTATGCC 865

Qy      280 GluThrIlePheGluGluAlaLysPheLysAspSerSerLeuAlaSerValValValGly 299
               :::|||||||||||||||||||||||||||||||||||||||||||||   ||||||
Db      866 AACAGCATCTTCGAGGAGGCCAAGTTCAAGGACAGCAGCCTGGCCTCGGTCACTGTGGGC 925

Qy      300 ValIleGlnValLeuPheThrAlaValAlaAlaLeuIleMetAspArgAlaGlyArgArg 319
            :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      926 ATAATCCAGGTCCTGTTCACTGCTGTGGCGGCCCTCATCATGGACAGAGCAGGGCGAAGG 985

Qy      320 LeuLeuLeuValLeuSerGlyValValMetValPheSerThrSerAlaPheGlyAlaTyr 339
            |||||||||   ||||||||||||:::||||||||||||   ||||||||||||   |||
Db      986 CTGCTCCTGGCCTTGTCGGGTGTGATCATGGTGTTCAGTATGAGTGCCTTTGGTACCTAC 1045

Qy      340 PheLysLeuThrGlnGlyGlyProGlyAsnSerSerHisValAlaIleSerAlaProVal 359
            |||||||||||||||      |||   |||||||||||||||   :::      |||:::
Db     1046 TTCAAACTGACCCAGAGCCTCCCCAGCAACTCCTCCCACGTAGGCCTG---GTGCCCATC 1102

Qy      360 SerAlaGlnProValAspAlaSerValGlyLeuAlaTrpLeuAlaValGlySerMetCys 379
            :::|||:::|||||||||      ||||||||||||||||||||||||||||||||||||
Db     1103 GCGGCGGAGCCTGTGGATGTCCAAGTGGGACTGGCCTGGCTGGCTGTAGGCAGCATGTGC 1162

Qy      380 LeuPheIleAlaGlyPheAlaValGlyTrpGlyProIleProTrpLeuLeuMetSerGlu 399
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1163 CTCTTCATTGCTGGCTTTGCGGTGGGCTGGGGACCCATCCCCTGGCTCCTCATGTCAGAG 1222

Qy      400 IlePheProLeuHisValLysGlyValAlaThrGlyIleCysValLeuThrAsnTrpLeu 419
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1223 ATCTTCCCTCTGCATGTCAAGGGTGTGGCTACCGGCATCTGTGTCCTCACCAACTGGTTC 1282

Qy      420 MetAlaPheLeuValThrLysGluPheSerSerLeuMetGluValLeuArgProTyrGly 439
            |||||||||||||||||||||||||||||||||:::||||||:::|||||||||||||||
Db     1283 ATGGCCTTTCTAGTGACCAAAGAGTTCAGCAGCGTCATGGAGATGCTCAGACCCTACGGT 1342

Qy      440 AlaPheTrpLeuAlaSerAlaPheCysIlePheSerValLeuPheThrLeuPheCysVal 459
            ||||||||||||   :::|||||||||      ||||||||||||||||||      |||
Db     1343 GCCTTCTGGCTCACCGCTGCCTTCTGCGCTCTCAGTGTCCTATTCACACTGACCGTTGTC 1402

Qy      460 ProGluThrLysGlyLysThrLeuGluGlnIleThrAlaHisPheGluGlyArg 477
            |||||||||||||||:::||||||||||||:::|||||||||||||||||||||
Db     1403 CCTGAGACTAAAGGCAGGACTCTGGAACAAGTCACAGCCCATTTCGAGGGACGA 1456

RESULT 9
AF232061
LOCUS       AF232061                1843 bp    mRNA    linear   ROD 23-JUN-2000
DEFINITION  Mus musculus glucose transporter GLUT8 mRNA, complete cds.
ACCESSION   AF232061
VERSION     AF232061.1  GI:8671757
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1843)
  AUTHORS   Carayannopoulos,M.O., Chi,M.M., Cui,Y., Pingsterhaus,J.M.,
            McKnight,R.A., Mueckler,M., Devaskar,S.U. and Moley,K.H.
  TITLE     GLUT8 is a glucose transporter responsible for insulin-stimulated
            glucose uptake in the blastocyst
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (13), 7313-7318 (2000)
  MEDLINE   20319023
   PUBMED   10860996
REFERENCE   2  (bases 1 to 1843)
  AUTHORS   Moley,K.H., Carayannopoulos,M.O. and Cui,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-FEB-2000) OB/Gyn, Washington University, 4566 Scott
            Ave, St. Louis, MO 63110, USA
FEATURES             Location/Qualifiers
     source          1. .1843
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="129"
                     /db_xref="taxon:10090"
     5'UTR           1. .48
     CDS             49. .1482
                     /codon_start=1
                     /product="glucose transporter GLUT8"
                     /protein_id="AAF78366.1"
                     /db_xref="GI:8671758"
                     /translation="MSPEDPQETQPLLRPPEARTPRGRRVFLASFAAALGPLSFGFAL
                     GYSSPAIPSLRRTAPPALRLGDNAASWFGAVVTLGAAAGGILGGWLLDRSGRKLSLLL
                     CTVPFVTGFAVITAARDVWMLLGGRLLTGLACGVASLVAPVYISEIAYPAVRGLLGSC
                     VQLMVVTGILLAYVAGWVLEWRWLAVLGCVPPTLMLLLMCYMPETPRFLLTQHQYQEA
                     MAALRFLWGSEEGWEEPPVGAEHQGFQLALLRRPGIYKPLIIGISLMVFQQLSGVNAI
                     MFYANSIFEEAKFKDSSLASVTVGIIQVLFTAVAALIMDRAGRRLLLALSGVIMVFSM
                     SAFGTYFKLTQSLPSNSSHVGLVPIAAEPVDVQVGLAWLAVGSMCLFIAGFAVGWGPI
                     PWLLMSEIFPLHVKGVATGICVLTNWFMAFLVTKEFSSVMEMLRPYGAFWLTAAFCAL
                     SVLFTLTVVPETKGRTLEQVTAHFEGR"
     3'UTR           1483. .1843
ORIGIN

Alignment Scores:
Pred. No.:               8.88e-146     Length:         1843
Score:                   2156.00       Matches:        411
Percent Similarity:      92.47%        Conservative:   31
Best Local Similarity:   85.98%        Mismatches:     34
Query Match:             87.75%        Indels:         2
```

```
DB:                         10           Gaps:          2

US-09-886-954A-1 (1-477) x AF232061 (1-843)

Qy        1 MetThrProGluAspProGluGluThrGlnProLeuLeuGlyProProGlyGlySerAla 20
            |||:::||||||||||||:::||||||||||||||||||   ||||||
Db       49 ATGAGTCCCGAGGACCCCCAGGAGACTCAGCCGCTATTGCGGCCACCGGAAGCCAGGACT 108

Qy       21 ProArgGlyArgArgValPheLeuAlaAlaPheAlaAlaAlaLeuGlyProLeuSerPhe 40
            |||||||||||||||||||||||||||:::||||||||||||||||||||||||||||||
Db      109 CCCCGCGGCCGCCGGGTCTTCCTCGCTTCCTTCGCCGCCGCTCTGGGACCCCTCAGCTTC 168

Qy       41 GlyPheAlaLeuGlyTyrSerSerProAlaIleProSerLeuGlnArgAlaAlaProPro 60
            ||||||||||||||||||||||||||||||||||||||||||:::|||   |||||||||
Db      169 GGCTTCGCGCTCGGCTACAGCTCCCCCGCCATCCCCAGCCTGCGGCGCACCGCACCCCCG 228

Qy       61 AlaProArgLeuAspAspAlaAlaAlaSerTrpPheGlyAlaValValThrLeuGlyAla 80
            |||   ||||||   |||   |||||||||||||||||||||||||||||||||||||||
Db      229 GCCCTGCGCCTCGGAGACAATGCGGCCTCCTGGTTCGGGGCCGTCGTGACCCTGGGCGCT 288

Qy       81 AlaAlaGlyGlyValLeuGlyGlyTrpLeuValAspArgAlaGlyArgLysLeuSerLeu 100
            ||||||||||||:::|||||||||||||||:::||||||:::||||||||||||||||||
Db      289 GCTGCAGGGGGCATACTGGGCGGCTGGCTCCTGGACCGTTCAGGGCGCAAGCTGAGCCTC 348

Qy      101 LeuLeuCysSerValProPheValAlaGlyPheAlaValIleThrAlaAlaGlnAspVal 120
            |||||||||:::||||||||||||   ||||||||||||||||||||||||:::||||||
Db      349 TTGCTCTGCACCGTGCCCTTCGTGACTGGCTTTGCTGTCATCACCGCGGCCCGGGATGTG 408

Qy      121 TrpMetLeuLeuGlyGlyArgLeuLeuThrGlyLeuAlaCysGlyValAlaSerLeuVal 140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      409 TGGATGCTGCTCGGAGGCCGCCTCCTCACCGGCCTTGCCTGCGGAGTCGCCTCACTAGTG 468

Qy      141 AlaProValTyrIleSerGluIleAlaTyrProAlaValArgGlyLeuLeuGlySerCys 160
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      469 GCACCGGTTTACATCTCGGAAATCGCCTACCCAGCTGTCCGAGGACTGCTCGGCTCCTGT 528

Qy      161 ValGlnLeuMetValValValGlyIleLeuLeuAlaTyrLeuAlaGlyTrpValLeuGlu 180
            ||||||||||||||||||   ||||||||||||||||||:::||||||||||||||||||
Db      529 GTGCAGCTGATGGTTGTCACTGGCATCCTCCTGGCCTATGTGGCAGGCTGGGTCCTAGAG 588

Qy      181 TrpArgTrpLeuAlaValLeuGlyCysValProProSerLeuMetLeuLeuLeuMetCys 200
            ||||||||||||||||||||||||||||||||||||:::|||||||||||||||||||||
Db      589 TGGCGCTGGCTGGCCGTGCTGGGCTGTGTGCCCCCCACCCTCATGCTGCTGCTCATGTGC 648

Qy      201 PheMetProGluThrProArgPheLeuLeuThrGlnHisArgArgGlnGluAlaMetAla 220
            :::||||||||||||||||||||||||||||||||||||:::   |||||||||||||||
Db      649 TACATGCCCGAGACCCCACGTTTTCTCCTCACTCAACACCAGTACCAGGAGGCCATGGCT 708

Qy      221 AlaLeuArgPheLeuTrpGlySerGluGlnGlyTrpGluAspProProIleGlyAlaGlu 240
            |||||||||||||||||||||||||||:::|||||||||:::||||||:::|||||||||
Db      709 GCCTTGCGCTTCCTGTGGGGCTCTGAGGAGGGCTGGGAAGAGCCCCCTGTTGGGGCTGAG 768

Qy      241 ---GlnSerPheHisLeuAlaLeuLeuArgGlnProGlyIleTyrLysProPheIleIle 259
               |||   |||   |||||||||||||||:::||||||||||||||||||   ||||||
Db      769 CACCAGGGCTTCCAGCTGGCCCTGCTGAGGCGCCCTGGCATCTACAAGCCCCTCATCATC 828

Qy      260 GlyValSerLeuMetAlaPheGlnGlnLeuSerGlyValAsnAlaValMetPheTyrAla 279
            |||:::|||||||||   |||||||||||||||||||||||||||:::||||||||||||
Db      829 GGCATTTCCCTCATGGTCTTCCAGCAGCTGTCAGGGGTCAATGCTATCATGTTCTATGCC 888

Qy      280 GluThrIlePheGluGluAlaLysPheLysAspSerSerLeuAlaSerValValValGly 299
               :::|||||||||||||||||||||||||||||||||||||||||||||   ||||||
Db      889 AACAGCATCTTCGAGGAGGCCAAGTTCAAGGACAGCAGCCTGGCCTCGGTCACTGTGGGC 948

Qy      300 ValIleGlnValLeuPheThrAlaValAlaAlaLeuIleMetAspArgAlaGlyArgArg 319
            :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      949 ATAATCCAGGTCCTGTTCACTGCTGTGGCGGCCCTCATCATGGACAGAGCAGGGCGAAGG 1008

Qy      320 LeuLeuLeuValLeuSerGlyValValMetValPheSerThrSerAlaPheGlyAlaTyr 339
            |||||||||   ||||||||||||:::||||||||||||   ||||||||||||   |||
Db     1009 CTGCTCCTGGCCTTGTCGGGTGTGATCATGGTGTTCAGTATGAGTGCCTTTGGTACCTAC 1068

Qy      340 PheLysLeuThrGlnGlyGlyProGlyAsnSerSerHisValAlaIleSerAlaProVal 359
            |||||||||||||||      |||   |||||||||||||||   :::      |||:::
Db     1069 TTCAAACTGACCCAGAGCCTCCCCAGCAACTCCTCCCACGTAGGCCTG---GTGCCCATC 1125

Qy      360 SerAlaGlnProValAspAlaSerValGlyLeuAlaTrpLeuAlaValGlySerMetCys 379
            :::|||:::|||||||||      ||||||||||||||||||||||||||||||||||||
Db     1126 GCGGCGGAGCCTGTGGATGTCCAAGTGGGACTGGCCTGGCTGGCTGTAGGCAGCATGTGC 1185

Qy      380 LeuPheIleAlaGlyPheAlaValGlyTrpGlyProIleProTrpLeuLeuMetSerGlu 399
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1186 CTCTTCATTGCTGGCTTTGCGGTGGGCTGGGGACCCATCCCCTGGCTCCTCATGTCAGAG 1245

Qy      400 IlePheProLeuHisValLysGlyValAlaThrGlyIleCysValLeuThrAsnTrpLeu 419
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1246 ATCTTCCCTCTGCATGTCAAGGGTGTGGCTACCGGCATCTGTGTCCTCACCAACTGGTTC 1305

Qy      420 MetAlaPheLeuValThrLysGluPheSerSerLeuMetGluValLeuArgProTyrGly 439
            |||||||||||||||||||||||||||||||||:::||||||:::|||||||||||||||
Db     1306 ATGGCCTTTCTAGTGACCAAAGAGTTCAGCAGCGTCATGGAGATGCTCAGACCCTACGGT 1365

Qy      440 AlaPheTrpLeuAlaSerAlaPheCysIlePheSerValLeuPheThrLeuPheCysVal 459
            ||||||||||||   :::|||||||||      ||||||||||||||||||      |||
Db     1366 GCCTTCTGGCTCACCGCTGCCTTCTGCGCTCTCAGTGTCCTATTCACACTGACCGTTGTC 1425

Qy      460 ProGluThrLysGlyLysThrLeuGluGlnIleThrAlaHisPheGluGlyArg 477
            |||||||||||||||:::||||||||||||:::|||||||||||||||||||||
Db     1426 CCTGAGACTAAAGGCAGGACTCTGGAACAAGTCACAGCCCATTTCGAGGGACGA 1479

RESULT 10
AX076671
LOCUS       AX076671                2072 bp    DNA     linear   PAT 06-FEB-2001
DEFINITION  Sequence 5 from Patent WO0104145.
ACCESSION   AX076671
VERSION     AX076671.1  GI:12711202
KEYWORDS    .
SOURCE      Mus sp.
  ORGANISM  Mus sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Thorens,B., Ibberson,M. and Uldry,M.
  TITLE     Glutx polypeptide family and nucleic acids encoding same
  JOURNAL   Patent: WO 0104145-A 5 18-JAN-2001;
            University of Lausanne (CH)
FEATURES             Location/Qualifiers
     source          1. .2072
                     /organism="Mus sp."
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:10095"
     CDS             21. .1454
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAC28497.1"
                     /db_xref="GI:12711203"
                     /db_xref="REMTREMBL:CAC28497"
                     /translation="MSPEDPQETQPLLRPPEARTPRGRRVFLASFAAALGPLNFGFAL
                     GYSSPAIPSLRRTAPPALRLGDNAASWFGAVVTLGAAAGGILGGWLLDRSGRKLSLLL
                     CTVPFVTGFAVITAARDVWMLLGGRLLTGLACGVASLVAPVYISEIAYPAVRGLLGSC
                     VQLMVVTGILLAYVAGWVLEWRWLAVLGCVPPTLMLLLMCYMPETPRFLLTQHQYQEA
                     MAALRFLWGSEEGWEEPPVGAEHQGFQLALLRRPGIYKPLIIGISLMVFQQLSGVNAI
                     MFYANSIFEEAKFKDSSLASVTVGIIQVLFTAVAALIMDRAGRRLLLALSGVIMVFSM
                     SAFGTYFKLTQSLPSNSSHVGLVPIAAEPVDVQVGLAWLAVGSMCLFIAGFAVGWGPI
                     PWLLMSEIFPLHVKGVATGICVLTNWFMAFLVTKEFNSVMEMLRPYGAFWLTAAFCAL
                     SVLFTLTVVPETKGRTLEQVTAHFEGR"
ORIGIN

Alignment Scores:
Pred. No.:              2.75e-145     Length:         2072
Score:                  2150.00       Matches:        409
Percent Similarity:     92.47%        Conservative:   33
Best Local Similarity:  85.56%        Mismatches:     34
Query Match:            87.51%        Indels:         2
DB:                     6             Gaps:           2
```

```
US-09-886-954A-1 (1-477) x AX076671 (1-2072)

Qy          1 MetThrProGluAspProGluGluThrGlnProLeuLeuGlyProProGlyGlySerAla 20
              |||:::||||||||||||:::||||||||||||||||||   ||||||
Db         21 ATGTCTCCCGAGGACCCCCAGGAGACGCAGCCGCTATTGCGGCCACCGGAAGCCAGGACT 80

Qy         21 ProArgGlyArgArgValPheLeuAlaAlaPheAlaAlaAlaLeuGlyProLeuSerPhe 40
              |||||||||||||||||||||||||||:::||||||||||||||||||||||||:::|||
Db         81 CCCCGCGGCCGCCGGGTCTTCCTCGCTTCCTTCGCCGCCGCTCTGGACCCCTCAACTTC 140

Qy         41 GlyPheAlaLeuGlyTyrSerSerProAlaIleProSerLeuGlnArgAlaAlaProPro 60
              ||||||||||||||||||||||||||||||||||||||||||:::|||   |||||||||
Db        141 GGCTTCGCGCTCGGCTACAGCTCCCCCGCCATCCCCAGCCTGCGGCGCACCGCACCCCCG 200

Qy         61 AlaProArgLeuAspAspAlaAlaAlaSerTrpPheGlyAlaValValThrLeuGlyAla 80
              |||   ||||||   |||   |||||||||||||||||||||||||||||||||||||||
Db        201 GCCCTGCGCCTCGGAGACAATGCGGCCTCCTGGTTCGGGGCCGTCGTGACCCTGGGCGCT 260

Qy         81 AlaAlaGlyGlyValLeuGlyGlyTrpLeuValAspArgAlaGlyArgLysLeuSerLeu 100
              ||||||||||||:::|||||||||||||||:::||||||:::||||||||||||||||||
Db        261 GCTGCAGGGGGCATACTGGGCGGCTGGCTCCTGGACCGTTCAGGGCGCAAGCTGAGCCTC 320

Qy        101 LeuLeuCysSerValProPheValAlaGlyPheAlaValIleThrAlaAlaGlnAspVal 120
              |||||||||:::||||||||||||   ||||||||||||||||||||||||:::||||||
Db        321 TTGCTCTGCACCGTGCCCTTCGTGACTGGCTTTGCTGTCATCACCGCGGCCCGGGATGTG 380

Qy        121 TrpMetLeuLeuGlyGlyArgLeuLeuThrGlyLeuAlaCysGlyValAlaSerLeuVal 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        381 TGGATGCTGCTCGGAGGCCGCCTCCTCACCGGCCTTGCCTGCGGAGTCGCCTCACTAGTG 440

Qy        141 AlaProValTyrIleSerGluIleAlaTyrProAlaValArgGlyLeuLeuGlySerCys 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        441 GCACCGGTTTACATCTCGGAAATCGCCTACCCAGCTGTCCGAGGACTGCTCGGCTCCTGT 500

Qy        161 ValGlnLeuMetValValValGlyIleLeuLeuAlaTyrLeuAlaGlyTrpValLeuGlu 180
              ||||||||||||||||||   ||||||||||||||||||:::||||||||||||||||||
Db        501 GTGCAGCTGATGGTTGTCACTGGCATCCTCCTGGCCTATGTGGCAGGCTGGGTCCTAGAG 560

Qy        181 TrpArgTrpLeuAlaValLeuGlyCysValProProSerLeuMetLeuLeuLeuMetCys 200
              ||||||||||||||||||||||||||||||||||||:::|||||||||||||||||||||
Db        561 TGGCGCTGGCTGGCCGTGCTGGGCTGTGTGCCCCCCACCCTCATGCTGCTGCTCATGTGC 620

Qy        201 PheMetProGluThrProArgPheLeuLeuThrGlnHisArgArgGlnGluAlaMetAla 220
              :::||||||||||||||||||||||||||||||||||||:::   |||||||||||||||
Db        621 TACATGCCCGAGACCCCACGTTTTCTCCTCACTCAACACCAGTACCAGGAGGCCATGGCT 680

Qy        221 AlaLeuArgPheLeuTrpGlySerGluGlnGlyTrpGluAspProProIleGlyAlaGlu 240
              |||||||||||||||||||||||||||:::|||||||||:::||||||:::|||||||||
Db        681 GCCTTGCGCTTCCTGTGGGGCTCTGAGGAGGGCTGGGAAGAGCCCCCTGTTGGGGCTGAG 740

Qy        241 ---GlnSerPheHisLeuAlaLeuLeuArgGlnProGlyIleTyrLysProPheIleIle 259
                 |||   |||   |||||||||||||||:::||||||||||||||||||   ||||||
Db        741 CACCAGGGCTTCCAGCTGGCCCTGCTGAGGCGCCCTGGCATCTACAAGCCCCTCATCATC 800

Qy        260 GlyValSerLeuMetAlaPheGlnGlnLeuSerGlyValAsnAlaValMetPheTyrAla 279
              |||:::|||||||||   |||||||||||||||||||||||||||:::||||||||||||
Db        801 GGCATTTCCCTCATGGTCTTCCAGCAGCTGTCAGGGGTCAATGCTATCATGTTCTATGCC 860

Qy        280 GluThrIlePheGluGluAlaLysPheLysAspSerSerLeuAlaSerValValValGly 299
                 :::|||||||||||||||||||||||||||||||||||||||||||||   ||||||
Db        861 AACAGCATCTTCGAGGAGGCCAAGTTCAAGGACAGCAGCCTGGCCTCGGTCACTGTGGGC 920

Qy        300 ValIleGlnValLeuPheThrAlaValAlaAlaLeuIleMetAspArgAlaGlyArgArg 319
              :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        921 ATAATCCAGGTCCTGTTCACTGCTGTGGCGGCCCTCATCATGGACAGAGCAGGGCGAAGG 980

Qy        320 LeuLeuLeuValLeuSerGlyValValMetValPheSerThrSerAlaPheGlyAlaTyr 339
              |||||||||   ||||||||||||:::||||||||||||   ||||||||||||   |||
Db        981 CTGCTCCTGGCCTTGTCGGGTGTGATCATGGTGTTCAGTATGAGTGCCTTTGGTACCTAC 1040

Qy        340 PheLysLeuThrGlnGlyGlyProGlyAsnSerSerHisValAlaIleSerAlaProVal 359
              |||||||||||||||      |||   |||||||||||||||   :::      |||:::
Db       1041 TTCAAACTGACCCAGAGCCTCCCCAGCAACTCCTCCCACGTAGGCCTG---GTGCCCATC 1097

Qy        360 SerAlaGlnProValAspAlaSerValGlyLeuAlaTrpLeuAlaValGlySerMetCys 379
              :::|||:::|||||||||      ||||||||||||||||||||||||||||||||||||
Db       1098 GCGGCGGAGCCTGTGGATGTCCAAGTGGGACTGGCCTGGCTGGCTGTAGGCAGCATGTGC 1157

Qy        380 LeuPheIleAlaGlyPheAlaValGlyTrpGlyProIleProTrpLeuLeuMetSerGlu 399
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1158 CTCTTCATTGCTGGCTTTGCGGTGGGCTGGGGACCCATCCCCTGGCTCCTCATGTCAGAG 1217

Qy        400 IlePheProLeuHisValLysGlyValAlaThrGlyIleCysValLeuThrAsnTrpLeu 419
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1218 ATCTTCCCTCTGCATGTCAAGGGTGTGGCTACCGGCATCTGTGTCCTCACCAACTGGTTC 1277

Qy        420 MetAlaPheLeuValThrLysGluPheSerSerLeuMetGluValLeuArgProTyrGly 439
              |||||||||||||||||||||||||||:::|||:::||||||:::|||||||||||||||
Db       1278 ATGGCCTTTCTAGTGACCAAAGAGTTCAACAGCGTCATGGAGATGCTCAGACCCTACGGT 1337

Qy        440 AlaPheTrpLeuAlaSerAlaPheCysIlePheSerValLeuPheThrLeuPheCysVal 459
              ||||||||||||   :::|||||||||      ||||||||||||||||||      |||
Db       1338 GCCTTCTGGCTCACCGCTGCCTTCTGCGCTCTCAGTGTCCTATTCACACTGACCGTTGTC 1397

Qy        460 ProGluThrLysGlyLysThrLeuGluGlnIleThrAlaHisPheGluGlyArg 477
              |||||||||||||||:::||||||||||||:::|||||||||||||||||||||
Db       1398 CCTGAGACTAAAGGCAGGACTCTGGAACAAGTCACAGCCCATTTCGAGGGACGA 1451

RESULT 11
MMU245936
LOCUS       MMU245936            2072 bp    mRNA    linear   ROD 18-FEB-2000
DEFINITION  Mus musculus mRNA for glucose transporter (GLUTX1 gene).
ACCESSION   AJ245936
VERSION     AJ245936.1  GI:7018606
KEYWORDS    glucose transporter; GLUTX1 gene.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Ibberson,M., Uldry,M. and Thorens,B.
  TITLE     GLUTX1, a novel mammalian glucose transporter expressed in the
            central nervous system and insulin-sensitive tissues
  JOURNAL   J. Biol. Chem. 275 (7), 4607-4612 (2000)
  MEDLINE   20138191
   PUBMED   10671487
REFERENCE   2  (bases 1 to 2072)
  AUTHORS   Ibberson,M.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-SEP-1999) Ibberson M.R., Institute of Pharmacology
            and Toxicology, University of Lausanne, Rue du Bugnon 27, Lausanne,
            CH-1005, SWITZERLAND
FEATURES             Location/Qualifiers
     source          1..2072
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
     gene            1..2072
                     /gene="GLUTX1"
     CDS             21..1454
                     /gene="GLUTX1"
                     /codon_start=1
                     /product="glucose transporter"
                     /protein_id="CAB75719.1"
                     /db_xref="GI:7018607"
                     /db_xref="GOA:Q9JIF3"
                     /db_xref="SWISS-PROT:Q9JIF3"
                     /translation="MSPEDPQETQPLLRPPEARTPRGRRVFLASFAAALGPLNFGFAL
                     GYSSPAIPSLRRTAPPALRLGDNAASWFGAVVTLGAAAGGILGGWLLDRSGRKLSLLL
                     CTVPFVTGFAVITAARDVWMLLGGRLLTGLACGVASLVAPVYISEIAYPAVRGLLGSC
                     VQLMVVTGILLAYVAGWVLEWRWLAVLGCVPPTLMLLLMCYMPETPRFLLTQHQYQEA
                     MAALRFLWGSEEGWEEPPVGAEHQGFQLALLRRPGIYKPLIIGISLMVFQQLSGVNAI
                     MFYANSIFEEAKFKDSSLASVTVGIIQVLFTAVAALIMDRAGRRLLLALSGVIMVFSM
                     SAFGTYFKLTQSLPSNSSHVGLVPIAAEPVDVQVGLAWLAVGSMCLFIAGFAVGWGPI
                     PWLLMSEIFPLHVKGVATGICVLTNWFMAFLVTKEFNSVMEMLRPYGAFWLTAAFCAL
                     SVLFTLTVVPETKGRTLEQVTAHFEGR"
```

```
ORIGIN

Alignment Scores:
Pred. No.:              2.75e-145      Length:         2072
Score:                  2150.00        Matches:        409
Percent Similarity:     92.47%         Conservative:   33
Best Local Similarity:  85.56%         Mismatches:     34
Query Match:            87.51%         Indels:         2
DB:                     10             Gaps:           2

US-09-886-954A-1 (1-477) x MMU245936 (1-2072)

Qy            1 MetThrProGluAspProGluGluThrGlnProLeuLeuGlyProProGlyGlySerAla 20
                |||:::||||||||||||:::||||||||||||||||||   ||||||
Db           21 ATGTCTCCCGAGGACCCCCAGGAGACGCAGCCGCTATTGCGGCCACCGGAAGCCAGGACT 80

Qy           21 ProArgGlyArgArgValPheLeuAlaAlaPheAlaAlaAlaLeuGlyProLeuSerPhe 40
                |||||||||||||||||||||||||||:::||||||||||||||||||||||||:::|||
Db           81 CCCCGCGGCCGCCGGGTCTTCCTCGCTTCCTTCGCCGCCGCTCTGGGACCCCTCAACTTC 140

Qy           41 GlyPheAlaLeuGlyTyrSerSerProAlaIleProSerLeuGlnArgAlaAlaProPro 60
                ||||||||||||||||||||||||||||||||||||||||||:::|||   |||||||||
Db          141 GGCTTCGCGCTCGGCTACAGCTCCCCCGCCATCCCCAGCCTGCGGCGCACCGCACCCCCG 200

Qy           61 AlaProArgLeuAspAspAlaAlaAlaSerTrpPheGlyAlaValValThrLeuGlyAla 80
                |||   ||||||   |||   |||||||||||||||||||||||||||||||||||||||
Db          201 GCCCTGCGCCTCGGAGACAATGCGGCCTCCTGGTTCGGGGCCGTCGTGACCCTGGGCGCT 260

Qy           81 AlaAlaGlyGlyValLeuGlyGlyTrpLeuValAspArgAlaGlyArgLysLeuSerLeu 100
                ||||||||||||:::|||||||||||||||:::||||||:::||||||||||||||||||
Db          261 GCTGCAGGGGGCATACTGGGCGGCTGGCTCCTGGACCGTTCAGGGCGCAAGCTGAGCCTC 320

Qy          101 LeuLeuCysSerValProPheValAlaGlyPheAlaValIleThrAlaAlaGlnAspVal 120
                |||||||||:::||||||||||||   ||||||||||||||||||||||||:::||||||
Db          321 TTGCTCTGCACCGTGCCCTTCGTGACTGGCTTTGCTGTCATCACCGCGGCCCGGGATGTG 380

Qy          121 TrpMetLeuLeuGlyGlyArgLeuLeuThrGlyLeuAlaCysGlyValAlaSerLeuVal 140
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          381 TGGATGCTGCTCGGAGGCCGCCTCCTCACCGGCCTTGCCTGCGGAGTCGCCTCACTAGTG 440

Qy          141 AlaProValTyrIleSerGluIleAlaTyrProAlaValArgGlyLeuLeuGlySerCys 160
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          441 GCACCGGTTTACATCTCGGAAATCGCCTACCCAGCTGTCCGAGGACTGCTCGGCTCCTGT 500

Qy          161 ValGlnLeuMetValValValGlyIleLeuLeuAlaTyrLeuAlaGlyTrpValLeuGlu 180
                ||||||||||||||||||   ||||||||||||||||||:::||||||||||||||||||
Db          501 GTGCAGCTGATGGTTGTCACTGGCATCCTCCTGGCCTATGTGGCAGGCTGGGTCCTAGAG 560

Qy          181 TrpArgTrpLeuAlaValLeuGlyCysValProProSerLeuMetLeuLeuLeuMetCys 200
                ||||||||||||||||||||||||||||||||||||:::|||||||||||||||||||||
Db          561 TGGCGCTGGCTGGCCGTGCTGGGCTGTGTGCCCCCCACCCTCATGCTGCTGCTCATGTGC 620

Qy          201 PheMetProGluThrProArgPheLeuLeuThrGlnHisArgArgGlnGluAlaMetAla 220
                :::||||||||||||||||||||||||||||||||||||:::   |||||||||||||||
Db          621 TACATGCCCGAGACCCCACGTTTTCTCCTCACTCAACACCAGTACCAGGAGGCCATGGCT 680

Qy          221 AlaLeuArgPheLeuTrpGlySerGluGlnGlyTrpGluAspProProIleGlyAlaGlu 240
                |||||||||||||||||||||||||||:::|||||||||:::||||||:::|||||||||
Db          681 GCCTTGCGCTTCCTGTGGGGCTCTGAGGAGGGCTGGGAAGAGCCCCCTGTTGGGGCTGAG 740

Qy          241 ---GlnSerPheHisLeuAlaLeuLeuArgGlnProGlyIleTyrLysProPheIleIle 259
                   |||   |||   |||||||||||||||:::||||||||||||||||||   ||||||
Db          741 CACCAGGGCTTCCAGCTGGCCCTGCTGAGGCGCCCTGGCATCTACAAGCCCCTCATCATC 800

Qy          260 GlyValSerLeuMetAlaPheGlnGlnLeuSerGlyValAsnAlaValMetPheTyrAla 279
                |||:::|||||||||   |||||||||||||||||||||||||||:::||||||||||||
Db          801 GGCATTTCCCTCATGGTCTTCCAGCAGCTGTCAGGGGTCAATGCTATCATGTTCTATGCC 860

Qy          280 GluThrIlePheGluGluAlaLysPheLysAspSerSerLeuAlaSerValValValGly 299
                   :::|||||||||||||||||||||||||||||||||||||||||||||   ||||||
Db          861 AACAGCATCTTCGAGGAGGCCAAGTTCAAGGACAGCAGCCTGGCCTCGGTCACTGTGGGC 920

Qy          300 ValIleGlnValLeuPheThrAlaValAlaAlaLeuIleMetAspArgAlaGlyArgArg 319
                :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          921 ATAATCCAGGTCCTGTTCACTGCTGTGGCGGCCCTCATCATGGACAGAGCAGGGCGAAGG 980

Qy          320 LeuLeuLeuValLeuSerGlyValValMetValPheSerThrSerAlaPheGlyAlaTyr 339
                |||||||||   ||||||||||||:::||||||||||||   ||||||||||||   |||
Db          981 CTGCTCCTGGCCTTGTCGGGTGTGATCATGGTGTTCAGTATGAGTGCCTTTGGTACCTAC 1040

Qy          340 PheLysLeuThrGlnGlyGlyProGlyAsnSerSerHisValAlaIleSerAlaProVal 359
                |||||||||||||||      |||   |||||||||||||||   :::      |||:::
Db         1041 TTCAAACTGACCCAGAGCCTCCCCAGCAACTCCTCCCACGTAGGCCTG---GTGCCCATC 1097

Qy          360 SerAlaGlnProValAspAlaSerValGlyLeuAlaTrpLeuAlaValGlySerMetCys 379
                :::|||:::|||||||||      ||||||||||||||||||||||||||||||||||||
Db         1098 GCGGCGGAGCCTGTGGATGTCCAAGTGGGACTGGCCTGGCTGGCTGTAGGCAGCATGTGC 1157

Qy          380 LeuPheIleAlaGlyPheAlaValGlyTrpGlyProIleProTrpLeuLeuMetSerGlu 399
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1158 CTCTTCATTGCTGGCTTTGCGGTGGGCTGGGGACCCATCCCCTGGCTCCTCATGTCAGAG 1217

Qy          400 IlePheProLeuHisValLysGlyValAlaThrGlyIleCysValLeuThrAsnTrpLeu 419
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1218 ATCTTCCCTCTGCATGTCAAGGGTGTGGCTACCGGCATCTGTGTCCTCACCAACTGGTTC 1277

Qy          420 MetAlaPheLeuValThrLysGluPheSerSerLeuMetGluValLeuArgProTyrGly 439
                |||||||||||||||||||||||||||:::|||:::||||||:::|||||||||||||||
Db         1278 ATGGCCTTTCTAGTGACCAAAGAGTTCAACAGCGTCATGGAGATGCTCAGACCCTACGGT 1337

Qy          440 AlaPheTrpLeuAlaSerAlaPheCysIlePheSerValLeuPheThrLeuPheCysVal 459
                ||||||||||||   :::|||||||||      ||||||||||||||||||      |||
Db         1338 GCCTTCTGGCTCACCGCTGCCTTCTGCGCTCTCAGTGTCCTATTCACACTGACCGTTGTC 1397

Qy          460 ProGluThrLysGlyLysThrLeuGluGlnIleThrAlaHisPheGluGlyArg 477
                |||||||||||||||:::||||||||||||:::|||||||||||||||||||||
Db         1398 CCTGAGACTAAAGGCAGGACTCTGGAACAAGTCACAGCCCATTTCGAGGGACGA 1451

RESULT 12
AB033418
LOCUS       AB033418                2189 bp    mRNA    linear   ROD 11-APR-2000
DEFINITION  Rattus norvegicus glut8 mRNA for glucose transporter 8, complete
            cds.
ACCESSION   AB033418
VERSION     AB033418.1  GI:7592743
KEYWORDS    glucose transporter 8.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 2189)
  AUTHORS   Ishibashi,K.
  TITLE     Molecular cloning of a new putative glucose transporter
  JOURNAL   Published Only in DataBase (2000)
REFERENCE   2  (bases 1 to 2189)
  AUTHORS   Ishibashi,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-OCT-1999) Kenichi Ishibashi, Jichi Medical School,
            Pharmacology; Minami-kawachi, Kawachi, Tochigi 329-0498, Japan
            (E-mail:kishiba@jichi.ac.jp, Tel:81-285-58-7326,
            Fax:81-285-44-5541)
FEATURES             Location/Qualifiers
     source          1..2189
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10116"
                     /clone_lib="testis"
     gene            1..2189
                     /gene="glut8"
     CDS             109..1548
                     /gene="glut8"
                     /note="Glut8 has 12 transmembrane segments"
                     /codon_start=1
                     /product="glucose transporter 8"
                     /protein_id="BAA94383.1"
```

```
                     /db_xref="GI:7592744"
                     /translation="MSPEDPQETQPLLRSPGARAPGGRRVFLATFAAALGPLSFGFAL
                     GYSSPAIPSLRRTAPPALRLGDTAASWFGAVVTLGAAAQGAVLGGWLLDRAGRKLSLL
                     LCTVPFVTGFAVITAARDVWMLLGGRLLTGLACGVASLVAPVYISEIAYPAVRGLLGS
                     CVQLMVVTGILLAYVAGWVLEWRWLAVLGCVPPTLMLLLMCYMPETPRFLLTQHQYQE
                     AMAALRFLWGSEEGWEEPPVGAEHQGFQLAMLRRPGVHKPLIIGICLMVFQQLSGVNA
                     IMFYANTIFEEAKFKDSSLASVTVGIIQVLFTAVAALIMDRAGRKLLLALSGVIMVFS
                     MSAFGTYFKLTQSGPSNSSHVGLLVPISAEPADVHLGLAWLAVGSMCLFIAGFAVGWG
                     PIPWLLMSEIFPLHIKGVATGVCVLTNWFMAFLVTKEFNSIMEILRPYGAFWLTAAFC
                     ILSVLFTLTFVPETKGRTLEQITAHFEGR"
ORIGIN

Alignment Scores:
Pred. No.:              5.7e-145        Length:          2189
Score:                  2146.00         Matches:         407
Percent Similarity:     91.86%          Conservative:    33
Best Local Similarity:  84.97%          Mismatches:      37
Query Match:            87.34%          Indels:          2
DB:                     10              Gaps:            2

US-09-886-954A-1 (1-477) x AB033418 (1-2189)

Qy           1 MetThrProGluAspProGluGluThrGlnProLeuLeuGlyProProGlyGlySerAla 20
               |||:::||||||||||||:::||||||||||||||||||      ||||||      |||
Db         109 ATGTCGCCCGAGGACCCCCAGGAAACACAGCCGCTGTTGCGGTCGCCGGGCGCCAGGGCT 168

Qy          21 ProArgGlyArgArgValPheLeuAlaAlaPheAlaAlaAlaLeuGlyProLeuSerPhe 40
               |||   |||||||||||||||||||||   ||||||||||||||||||||||||||||||
Db         169 CCCGGCGGCCGCCGGGTCTTCCTTGCCACCTTCGCCGCCGCCCTGGGCCCGCTCAGCTTC 228

Qy          41 GlyPheAlaLeuGlyTyrSerSerProAlaIleProSerLeuGlnArgAlaAlaProPro 60
               ||||||||||||||||||||||||||||||||||||||||||:::|||   |||||||||
Db         229 GGCTTCGCGCTTGGCTACAGCTCCCCCGCCATCCCCAGCCTGCGGCGCACCGCACCCCCT 288

Qy          61 AlaProArgLeuAspAspAlaAlaAlaSerTrpPheGlyAlaValValThrLeuGlyAla 80
               |||   ||||||   |||   |||||||||||||||||||||||||||||||||||||||
Db         289 GCCCTACGCCTCGGAGACACTGCGGCCTCCTGGTTCGGGGCCGTCGTGACCCTGGGCGCT 348

Qy          81 AlaAla---GlyGlyValLeuGlyGlyTrpLeuValAspArgAlaGlyArgLysLeuSer 99
               ||||||   |||   ||||||||||||||||||:::||||||||||||||||||||||||
Db         349 GCGGCCCAGGGGGCCGTGCTGGGCGGCTGGCTCCTGGACCGTGCAGGGCGCAAGCTGAGC 408

Qy         100 LeuLeuLeuCysSerValProPheValAlaGlyPheAlaValIleThrAlaAlaGlnAsp 119
               ||||||||||||:::||||||||||||   ||||||||||||||||||||||||:::|||
Db         409 CTCCTGCTCTGCACCGTGCCCTTCGTGACCGGTTTTGCTGTCATCACCGCGGCCCGGGAT 468

Qy         120 ValTrpMetLeuLeuGlyGlyArgLeuLeuThrGlyLeuAlaCysGlyValAlaSerLeu 139
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         469 GTGTGGATGCTGCTCGGAGGCCGCCTCCTCACCGGCCTAGCCTGCGGAGTCGCCTCATTA 528

Qy         140 ValAlaProValTyrIleSerGluIleAlaTyrProAlaValArgGlyLeuLeuGlySer 159
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         529 GTGGCACCGGTCTATATCTCGGAAATCGCCTACCCAGCCGTTCGAGGACTGCTCGGCTCC 588

Qy         160 CysValGlnLeuMetValValValGlyIleLeuLeuAlaTyrLeuAlaGlyTrpValLeu 179
               |||||||||||||||||||||   ||||||||||||||||||:::|||||||||||||||
Db         589 TGTGTGCAGCTGATGGTTGTCACTGGCATCCTCTTGGCCTACGTGGCAGGCTGGGTCCTA 648

Qy         180 GluTrpArgTrpLeuAlaValLeuGlyCysValProProSerLeuMetLeuLeuLeuMet 199
               |||||||||||||||||||||||||||||||||||||||:::||||||||||||||||||
Db         649 GAATGGCGCTGGTTGGCTGTGCTGGGCTGTGTGCCCCCCACCCTCATGCTGCTGCTCATG 708

Qy         200 CysPheMetProGluThrProArgPheLeuLeuThrGlnHisArgArgGlnGluAlaMet 219
               |||:::||||||||||||||||||||||||||||||||||||:::   ||||||||||||
Db         709 TGCTACATGCCCGAGACCCCACGCTTCCTCCTGACGCAACACCAGTACCAGGAGGCCATG 768

Qy         220 AlaAlaLeuArgPheLeuTrpGlySerGluGlnGlyTrpGluAspProProIleGlyAla 239
               ||||||||||||||||||||||||||||||:::|||||||||:::||||||:::||||||
Db         769 GCTGCCCTGCGCTTCCTGTGGGGCTCTGAGGAGGGCTGGGAAGAGCCCCCTGTTGGGGCT 828

Qy         240 Glu---GlnSerPheHisLeuAlaLeuLeuArgGlnProGlyIleTyrLysProPheIle 258
               |||   |||   |||   ||||||:::||||||:::||||||::::::||||||   |||
Db         829 GAGCACCAGGGCTTCCAGCTGGCTATGCTGAGGCGCCCTGGTGTCCACAAGCCCCTCATC 888

Qy         259 IleGlyValSerLeuMetAlaPheGlnGlnLeuSerGlyValAsnAlaValMetPheTyr 278
               ||||||:::   ||||||   |||||||||||||||||||||||||||:::|||||||||
Db         889 ATCGGCATTTGCCTCATGGTCTTCCAGCAGCTGTCAGGGGTCAACGCCATCATGTTCTAT 948

Qy         279 AlaGluThrIlePheGluGluAlaLysPheLysAspSerSerLeuAlaSerValValVal 298
               |||   ||||||||||||||||||||||||||||||||||||||||||||||||   |||
Db         949 GCCAACACCATCTTTGAGGAGGCCAAGTTCAAGGACAGCAGCCTGGCCTCGGTCACTGTG 1008

Qy         299 GlyValIleGlnValLeuPheThrAlaValAlaAlaLeuIleMetAspArgAlaGlyArg 318
               |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1009 GGCATCATCCAGGTTCTGTTCACTGCTGTGGCGGCCCTCATCATGGACAGAGCAGGGCGA 1068

Qy         319 ArgLeuLeuLeuValLeuSerGlyValValMetValPheSerThrSerAlaPheGlyAla 338
               :::|||||||||   ||||||||||||:::||||||||||||   ||||||||||||
Db        1069 AAACTGCTTCTGGCCTTGTCGGGTGTGATCATGGTTTTCAGCATGAGCGCCTTTGGTACC 1128

Qy         339 TyrPheLysLeuThrGlnGlyGlyProGlyAsnSerSerHisValAlaIleSerAlaPro 358
               ||||||||||||||||||   ||||||   |||||||||||||||   :::      |||
Db        1129 TACTTCAAACTGACCCAGAGTGGCCCCAGCAACTCCTCCCATGTAGGCCTCCTGGTGCCC 1188

Qy         359 ValSerAlaGlnProValAspAlaSerValGlyLeuAlaTrpLeuAlaValGlySerMet 378
               :::||||||:::|||   |||      :::||||||||||||||||||||||||||||||
Db        1189 ATCTCCGCAGAGCCTGCTGATGTTCACCTGGGGCTGGCCTGGCTGGCTGTAGGCAGCATG 1248

Qy         379 CysLeuPheIleAlaGlyPheAlaValGlyTrpGlyProIleProTrpLeuLeuMetSer 398
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1249 TGCCTCTTCATCGCTGGTTTTGCAGTAGGCTGGGGACCCATCCCCTGGCTCCTCATGTCA 1308

Qy         399 GluIlePheProLeuHisValLysGlyValAlaThrGlyIleCysValLeuThrAsnTrp 418
               ||||||||||||||||||:::||||||||||||||||||:::||||||||||||||||||
Db        1309 GAGATCTTCCCTCTGCACATCAAGGGTGTGGCTACCGGCGTCTGTGTCCTCACCAACTGG 1368

Qy         419 LeuMetAlaPheLeuValThrLysGluPheSerSerLeuMetGluValLeuArgProTyr 438
                  |||||||||||||||||||||||||||:::|||:::||||||:::||||||||||||
Db        1369 TTCATGGCCTTTCTGGTGACCAAAGAGTTTAACAGCATCATGGAGATCCTCAGACCCTAC 1428

Qy         439 GlyAlaPheTrpLeuAlaSerAlaPheCysIlePheSerValLeuPheThrLeuPheCys 458
               |||||||||||||||   :::||||||||||||   ||||||||||||||||||
Db        1429 GGCGCCTTCTGGCTCACCGCTGCCTTCTGTATCCTCAGCGTCCTTTTCACGCTCACCTTT 1488

Qy         459 ValProGluThrLysGlyLysThrLeuGluGlnIleThrAlaHisPheGluGlyArg 477
               ||||||||||||||||||:::||||||||||||||||||||||||||||||||||||
Db        1489 GTCCCTGAGACTAAAGGCAGGACTCTGGAACAAATCACAGCCCATTTCGAGGGACGG 1545

RESULT 13
AX191507
LOCUS       AX191507                1461 bp    DNA     linear   PAT 15-AUG-2001
DEFINITION  Sequence 29 from Patent WO0149728.
ACCESSION   AX191507
VERSION     AX191507.1  GI:15209697
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Kato,S. and Kimura,T.
  TITLE     Human proteins having hydrophobic domains and dnas encoding these
            proteins
  JOURNAL   Patent: WO 0149728-A 29 12-JUL-2001;
            Protegene Inc. (JP) ; SAGAMI CHEMICAL RESEARCH CENTER (JP)
FEATURES             Location/Qualifiers
     source          1..1461
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
     CDS             61..849
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAC51153.1"
                     /db_xref="GI:15209698"
                     /db_xref="REMTREMBL:CAC51153"
```

```
                     /translation="MTPEDPEETQPLLGPPGGSAPRGRRVFLAAFAAALGPLSFGFAL
                     GYSSPAIPSLQRAAPPAPRLDDAAASWFGAVVTLGAAAGGVLGGWLVDRAGRKLSLLL
                     CSVPFVAGFAVITAAQDVWMLLGGRLLTGLACGVASLVAPVYISEIAYPAVRGLLGSC
                     VQLMVVVGILLAYLAGWVLEWRWLAVLGCVPPSLMLLLMCFMPETPRFLLTQHRRQEA
                     APGLVRCGHGVQHECLRRLLQADPGWPWQLLARGHLGACLCTAC"
ORIGIN

Alignment Scores:
Pred. No.:              1.09e-102      Length:        1461
Score:                  1554.50        Matches:       324
Percent Similarity:     67.85%         Conservative:  1
Best Local Similarity:  67.64%         Mismatches:    1
Query Match:            63.27%         Indels:        153
DB:                     6              Gaps:          2

US-09-886-954A-1 (1-477) x AX191507 (1-1461)

Qy          1 MetThrProGluAspProGluGluThrGlnProLeuLeuGlyProProGlyGlySerAla 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ATGACGCCCGAGGACCCAGAGGAAACCCAGCCGCTTCTGGGGCCTCCTGGCGGCAGCGCG 120

Qy         21 ProArgGlyArgArgValPheLeuAlaAlaPheAlaAlaAlaLeuGlyProLeuSerPhe 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCCCGCGGCCGCCGCGTCTTCCTCGCCGCCTTCGCCGCTGCCCTGGGCCCACTCAGCTTC 180

Qy         41 GlyPheAlaLeuGlyTyrSerSerProAlaIleProSerLeuGlnArgAlaAlaProPro 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCTTCGCGCTCGGCTACAGCTCCCCGGCCATCCCTAGCCTGCAGCGCGCCGCGCCCCCG 240

Qy         61 AlaProArgLeuAspAspAlaAlaAlaSerTrpPheGlyAlaValValThrLeuGlyAla 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCCCCGCGCCTGGACGACGCCGCCGCCTCCTGGTTCGGGGCTGTCGTGACCCTGGGTGCC 300

Qy         81 AlaAlaGlyGlyValLeuGlyGlyTrpLeuValAspArgAlaGlyArgLysLeuSerLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCGGCGGGGGGAGTGCTGGGCGGCTGGCTGGTGGACCGCGCCGGGCGCAAGCTGAGCCTC 360

Qy        101 LeuLeuCysSerValProPheValAlaGlyPheAlaValIleThrAlaAlaGlnAspVal 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TTGCTGTGCTCCGTGCCCTTCGTGGCCGGCTTTGCCGTCATCACCGCGGCCCAGGACGTG 420

Qy        121 TrpMetLeuLeuGlyGlyArgLeuLeuThrGlyLeuAlaCysGlyValAlaSerLeuVal 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TGGATGCTGCTGGGGGGCCGCCTCCTCACCGGCCTGGCCTGCGGTGTTGCCTCCCTAGTG 480

Qy        141 AlaProValTyrIleSerGluIleAlaTyrProAlaValArgGlyLeuLeuGlySerCys 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GCCCCGGTCTACATCTCCGAAATCGCCTACCCAGCAGTCCGGGGGTTGCTCGGCTCCTGT 540

Qy        161 ValGlnLeuMetValValValGlyIleLeuLeuAlaTyrLeuAlaGlyTrpValLeuGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTGCAGCTAATGGTCGTCGTCGGCATCCTCCTGGCCTACCTGGCAGGCTGGGTGCTGGAG 600

Qy        181 TrpArgTrpLeuAlaValLeuGlyCysValProProSerLeuMetLeuLeuLeuMetCys 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TGGCGCTGGCTGGCTGTGCTGGGCTGCGTGCCCCCCTCCCTCATGCTGCTTCTCATGTGC 660

Qy        201 PheMetProGluThrProArgPheLeuLeuThrGlnHisArgArgGlnGluAla-MetAl 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||  :::
Db        661 TTCATGCCCGAGACCCCGCGCTTCCTGCTGACTCAGCACAGGCGCCAGGAGGCTGCTC-- 718

Qy        220 aAlaLeuArgPheLeuTrpGlySerGluGlnGlyTrpGluAspProProIleGlyAlaGl 240

Db        718 ------------------------------------------------------------ 718

Qy        240 uGlnSerPheHisLeuAlaLeuLeuArgGlnProGlyIleTyrLysProPheIleIleGl 260

Db        718 ------------------------------------------------------------ 718

Qy        260 yValSerLeuMetAlaPheGlnGlnLeuSerGlyValAsnAlaValMetPheTyrAlaGl 280

Db        718 ------------------------------------------------------------ 718

Qy        280 uThrIlePheGluGluAlaLysPheLysAspSerSerLeuAlaSerValValValGlyVa 300

Db        718 ------------------------------------------------------------ 718

Qy        300 lIleGlnValLeuPheThrAlaValAlaAlaLeuIleMetAspArgAlaGlyArgArgLe 320

Db        718 ------------------------------------------------------------ 718

Qy        320 uLeuLeuValLeuSerGlyValValMetValPheSerThrSerAlaPheGlyAlaTyrPh 340
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        719 ----CTGGTCTTGTCAGGTGTGGTCATGGTGTTCAGCACGAGTGCCTTCGGCGCCTACTT 774

Qy        340 eLysLeuThrGlnGlyGlyProGlyAsnSerSerHisValAlaIleSerAlaProValSe 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        775 CAAGCTGACCCAGGGTGGCCCTGGCAACTCCTCGCACGTGGCCATCTCGGCGCCTGTCTC 834

Qy        360 rAlaGlnProValAspAlaSerValGlyLeuAlaTrpLeuAlaValGlySerMetCysLe 380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        835 TGCACAGCCTGTTGATGCCAGCGTGGGGCTGGCCTGGCTGGCCGTGGGCAGCATGTGCCT 894

Qy        380 uPheIleAlaGly-PheAlaValGlyTrpGlyProIleProTrpLeuLeuMetSerGluI 400
              |||||||||||||
Db        895 CTTCATCGCCGGAGGTC------------------------------------------- 911

Qy        400 lePheProLeuHisValLysGlyValAlaThrGlyIleCysValLeuThrAsnTrpLeuM 420

Db        911 ------------------------------------------------------------ 911

Qy        420 etAlaPheLeuValThrLysGluPheSerSerLeuMetGluValLeuArgProTyrGlyA 440
                                                          ||||||||||||||||
Db        912 --------------------------------------------CTCAGGCCCTATGGAG 927

Qy        440 laPheTrpLeuAlaSerAlaPheCysIlePheSerValLeuPheThrLeuPheCysValP 460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        928 CCTTCTGGCTTGCCTCCGCTTTCTGCATCTTCAGTGTCCTTTTCACTTTGTTCTGTGTCC 987

Qy        460 roGluThrLysGlyLysThrLeuGluGlnIleThrAlaHisPheGluGlyArg 477
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        988 CTGAAACTAAAGGAAAGACTCTGGAACAAATCACAGCCCATTTTGAGGGGCGA 1040

RESULT 14
AF321324
LOCUS       AF321324                1012 bp    mRNA    linear   MAM 02-JUL-2001
DEFINITION  Bos taurus glucose transporter 8 (Glut8) mRNA, partial cds.
ACCESSION   AF321324
VERSION     AF321324.1  GI:14582715
KEYWORDS    .
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 1012)
  AUTHORS   Augustin,R., Navarrete-Santos,A. and Fischer,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-NOV-2000) Anatomy and Cell Biology, Faculty of
            Medicine, Martin Luther University Halle, Grosse Steinstrasse 52,
            Halle 061097, Germany
FEATURES             Location/Qualifiers
     source          1. .1012
                     /organism="Bos taurus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9913"
     gene            <1. .1012
                     /gene="Glut8"
     CDS             <1. .1005
                     /gene="Glut8"
                     /codon_start=1
                     /product="glucose transporter 8"
                     /protein_id="AAK69606.1"
                     /db_xref="GI:14582716"
                     /translation="ISEIAYPEVRGLLGSCVQLMVVTGILLAYLAGWVLEWRWLAVLG
                     CVAPSFMLLLMCFMPETPRFLLSQHKHQEAMAAMQFLWGYAQGWEEPPLGAQHQDFHV
```

```
                   AQLRRPGVYKPFIIGISLMAFQQLSGVNAVMFYAETIFEEAKFKDSSLASVVVGVIQV
                   LFTATAALIMDRAGRRLLLTLSGVVMVFSTSAFGTYFKLTEGGPSNSSHVDLPALVSM
                   EAADTNVGLAWLAVGNMCLFIAGFAVGWGPIPWLLMSEIFPLHVKGVATGVCVLTNWF
                   MAFLVTKEFSSLMEVLRPYGAFWLASAFCIFGVLFTLACVPETKGKTLEQITAHFEGR
                   "
ORIGIN

Alignment Scores:
Pred. No.:              1.17e-101      Length:         1012
Score:                  1537.50        Matches:        293
Percent Similarity:     92.81%         Conservative:   17
Best Local Similarity:  87.72%         Mismatches:     23
Query Match:            62.58%         Indels:         1
DB:                     4              Gaps:           1

US-09-886-954A-1 (1-477) x AF321324 (1-1012)

Qy        145 IleSerGluIleAlaTyrProAlaValArgGlyLeuLeuGlySerCysValGlnLeuMet 164
              |||||||||||||||||||||   ||||||||||||||||||||||||||||||||||||
Db          1 ATCTCGGAAATCGCCTACCCTGAGGTGCGAGGGCTGCTCGGCTCCTGTGTGCAGCTGATG 60

Qy        165 ValValValGlyIleLeuLeuAlaTyrLeuAlaGlyTrpValLeuGluTrpArgTrpLeu 184
              ||||||   |||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GTAGTCACAGGCATCCTCCTAGCCTACCTGGCAGGCTGGGTGCTCGAGTGGCGCTGGCTG 120

Qy        185 AlaValLeuGlyCysValProProSerLeuMetLeuLeuLeuMetCysPheMetProGlu 204
              ||||||||||||||||||   ||||||   ||||||||||||||||||||||||||||||
Db        121 GCCGTGCTGGGCTGTGTGGCCCCCTCCTTCATGCTGCTGCTCATGTGTTTCATGCCTGAG 180

Qy        205 ThrProArgPheLeuLeuThrGlnHisArgArgGlnGluAlaMetAlaAlaLeuArgPhe 224
              ||||||||||||||||||:::||||||:::   ||||||||||||||||||::::::|||
Db        181 ACCCCTCGCTTCCTGCTGTCTCAGCACAAGCACCAGGAGGCCATGGCCGCCATGCAGTTC 240

Qy        225 LeuTrpGlySerGluGlnGlyTrpGluAspProProIleGlyAlaGlu---GlnSerPhe 243
              |||||||||      ||||||||||||:::||||||:::||||||:::   |||   |||
Db        241 CTGTGGGGCTATGCCCAGGGTTGGGAGGAGCCTCCCCTTGGGGCTCAGCACCAGGACTTC 300

Qy        244 HisLeuAlaLeuLeuArgGlnProGlyIleTyrLysProPheIleIleGlyValSerLeu 263
              |||:::|||   ||||||:::||||||:::|||||||||||||||||||||:::||||||
Db        301 CACGTGGCCCAGCTGCGGCGTCCTGGTGTCTACAAGCCCTTCATCATCGGCATCTCCCTG 360

Qy        264 MetAlaPheGlnGlnLeuSerGlyValAsnAlaValMetPheTyrAlaGluThrIlePhe 283
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ATGGCCTTCCAGCAGCTGTCAGGCGTCAATGCCGTCATGTTCTACGCGGAGACCATCTTT 420

Qy        284 GluGluAlaLysPheLysAspSerSerLeuAlaSerValValValGlyValIleGlnVal 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GAGGAGGCTAAGTTCAAGGACAGCAGCCTGGCCTCAGTCGTCGTGGGCGTCATCCAGGTG 480

Qy        304 LeuPheThrAlaValAlaAlaLeuIleMetAspArgAlaGlyArgArgLeuLeuLeuVal 323
              ||||||||||||   ||||||||||||||||||||||||||||||||||||||||||
Db        481 CTGTTCACTGCCACGGCGGCCCTGATCATGGACAGAGCTGGGCGAAGGCTGCTCTTGACC 540

Qy        324 LeuSerGlyValValMetValPheSerThrSerAlaPheGlyAlaTyrPheLysLeuThr 343
              ||||||||||||||||||||||||||||||||||||||||||   |||||||||||||||
Db        541 TTGTCAGGTGTGGTCATGGTGTTCAGCACCAGCGCCTTTGGTACCTACTTCAAGCTGACC 600

Qy        344 GlnGlyGlyProGlyAsnSerSerHisValAlaIleSerAlaProValSerAlaGlnPro 363
              :::|||||||||   |||||||||||||||   :::   |||   ||||||   :::
Db        601 GAGGGCGGCCCCAGCAACTCCTCCCACGTGGACCTCCCGGCGCTCGTCTCCATGGAGGCC 660

Qy        364 ValAspAlaSerValGlyLeuAlaTrpLeuAlaValGlySerMetCysLeuPheIleAla 383
                 |||   :::|||||||||||||||||||||||||||:::||||||||||||||||||
Db        661 GCCGATACCAACGTGGGGCTGGCCTGGCTGGCGGTGGGCAACATGTGCCTCTTCATCGCC 720

Qy        384 GlyPheAlaValGlyTrpGlyProIleProTrpLeuLeuMetSerGluIlePheProLeu 403
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GGCTTCGCTGTGGGCTGGGGGCCCATCCCGTGGCTCCTCATGTCAGAGATCTTCCCTCTG 780

Qy        404 HisValLysGlyValAlaThrGlyIleCysValLeuThrAsnTrpLeuMetAlaPheLeu 423
              ||||||||||||||||||||||||:::||||||||||||||||||   ||||||||||||
Db        781 CATGTCAAGGGCGTGGCCACCGGCGTCTGCGTCCTCACCAACTGGTTCATGGCCTTTCTG 840

Qy        424 ValThrLysGluPheSerSerLeuMetGluValLeuArgProTyrGlyAlaPheTrpLeu 443
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GTGACCAAAGAGTTCAGCAGTCTCATGGAGGTGCTCAGGCCCTACGGTGCCTTCTGGCTG 900

Qy        444 AlaSerAlaPheCysIlePheSerValLeuPheThrLeuPheCysValProGluThrLys 463
              |||||||||||||||||||||   |||||||||||||||   ||||||||||||||||||
Db        901 GCCTCTGCCTTCTGCATCTTCGGTGTCCTTTTCACTCTGGCCTGTGTCCCTGAGACCAAA 960

Qy        464 GlyLysThrLeuGluGlnIleThrAlaHisPheGluGlyArg 477
              ||||||||||||||||||||||||||||||||||||||||||
Db        961 GGGAAGACTCTGGAACAAATCACAGCCCATTTTGAGGGCCGA 1002

RESULT 15
BC049409
LOCUS       BC049409                1929 bp    mRNA    linear   VRT 07-OCT-2003
DEFINITION  Danio rerio cDNA clone MGC:56364 IMAGE:5604396, complete cds.
ACCESSION   BC049409
VERSION     BC049409.1  GI:29436433
KEYWORDS    MGC.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 1929)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
  MEDLINE   22388257
   PUBMED   12477932
REFERENCE   2  (bases 1 to 1929)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-MAR-2003) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Sumio Sugano
            cDNA Library Preparation: Dr. Sumio Sugano
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Genome Sequence Centre,
            BC Cancer Agency, Vancouver, BC, Canada
            info@bcgsc.bc.ca
            Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
            Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
            Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
            Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
            Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline
            Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott,
            Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
            George Yang, Scott Zuyderduyn, Marco Marra.
```

```
            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 104 Row: b Column: 18
            This clone was selected for full length sequencing because it
            passed the following selection criteria: Hexamer frequency ORF
            analysis, Similarity but not identity to protein.
FEATURES             Location/Qualifiers
     source          1..1929
                     /organism="Danio rerio"
                     /mol_type="mRNA"
                     /db_xref="taxon:7955"
                     /clone="MGC:56364 IMAGE:5604396"
                     /tissue_type="Whole body, adult male"
                     /clone_lib="Sugano SJD adult male"
                     /lab_host="DH10B"
                     /note="Vector: pME18S-FL3"
     CDS             149..1645
                     /codon_start=1
                     /product="Unknown (protein for MGC:56364)"
                     /protein_id="AAH49409.1"
                     /db_xref="GI:29436434"
                     /translation="MNEYEENRPLLVNDERTEEDRSEQDAYLDKVKNGKLFIATFAAV
                     LGPLSFGFVLGYSSPAIPELRRIQDLRLQLSVEEASWFGSVVTIGAALGGLLGGWIVE
                     RIGRKLSLMFCAIPFIFGFTTIIAAQNHWMFYVGRVLTGLASGVTSLVVPLYISEMAH
                     ERVRGTLGSCVQLMVVIGIMGAYVTGLFLDWRWLAVASSIPPTLMLLSMCFMPETPRF
                     LLCQGKRREAEDALRFLRGPDAPAEWECARIEDAYKNEEQSFSLGDLKDPGVYKPLGI
                     GVMMMLLQQFTGINAIMFYAETIFEQAHFKSSDVATVIVAATQVVFTAIAALIMDKAG
                     RKVLLILSGVVMCVSEAVFGVYFKLTVMKPNNSSMTSVLTDTHGLLEDQPSADLAWLA
                     VGSMGFFIAGFAIGWGPTPWLVMSEIFPTRVRGLGSALCVLTNWTCAFIVTKTFQNLM
                     DALSSAGTFWMFSALCASNVVFTAFFVPETKGKTLEEIQAGFKGTRMR"
     misc_feature    257..1624
                     /note="sugar_tr; Region: Sugar (and other) transporter"
                     /db_xref="CDD:pfam00083"
ORIGIN

Alignment Scores:
Pred. No.:              2.85e-92       Length:          1929
Score:                  1411.50        Matches:         282
Percent Similarity:     70.99%         Conservative:    63
Best Local Similarity:  58.02%         Mismatches:      122
Query Match:            57.45%         Indels:          19
DB:                     5              Gaps:            5

US-09-886-954A-1 (1-477) x BC049409 (1-1929)

Qy         2 ThrProGluAspProGluGluThrGlnProLeuLeuGlyProProGlyGlySerAlaPro 21
             |||   ||||||      |||            |||
Db       197 ACAGAAGAGGATAGATCCGAGCAAGACGCATACCTG---------------GATAAAGTT 241

Qy        22 ArgGlyArgArgValPheLeuAlaAlaPheAlaAlaAlaLeuGlyProLeuSerPheGly 41
             :::      ::::::|||:::|||   |||||||||   |||||||||||||||||||||
Db       242 AAGAATGGAAAGCTGTTCATCGCCACCTTCGCAGCAGTTCTGGGACCGCTTAGTTTTGGG 301

Qy        42 PheAlaLeuGlyTyrSerSerProAlaIleProSerLeuGlnArgAlaAlaProProAla 61
             |||   |||||||||||||||||||||||||||   |||:::|||
Db       302 TTCGTGCTGGGCTACAGCTCTCCTGCTATCCCTGAACTCCGCAGAATCCAAGACCTGAGA 361

Qy        62 ProArgLeuAspAspAlaAlaAlaSerTrpPheGlyAlaValValThrLeuGlyAlaAla 81
                :::|||            |||||||||||||||:::|||||||||:::|||||||||
Db       362 CTGCAGCTCAGCGTAGAGGAAGCTTCCTGGTTCGGGTCTGTGGTGACCATTGGTGCGGCT 421

Qy        82 AlaGlyGlyValLeuGlyGlyTrpLeuValAspArgAlaGlyArgLysLeuSerLeuLeu 101
                ||||||:::||||||||||||:::|||:::|||   ||||||||||||||||||:::
Db       422 CTTGGGGGTCTTCTGGGAGGATGGATTGTGGAGAGGATCGGACGAAAGTTGAGCCTTATG 481

Qy       102 LeuCysSerValProPheValAlaGlyPheAlaValIleThrAlaAlaGlnAspValTrp 121
                |||::::::||||||:::   ||||||      |||   |||||||||:::   |||
Db       482 TTCTGTGCAATACCATTCATCTTTGGGTTTACCACCATTATCGCTGCTCAGAACCACTGG 541

Qy       122 MetLeuLeuGlyGlyArgLeuLeuThrGlyLeuAlaCysGlyValAlaSerLeuValAla 141
             |||         ||||||:::|||||||||||||||   ||||||   |||||||||
Db       542 ATGTTCTATGTGGGCAGAGTGCTCACTGGTTTGGCCAGTGGAGTGACCTCTCTAGTTGTA 601

Qy       142 ProValTyrIleSerGluIleAlaTyrProAlaValArgGlyLeuLeuGlySerCysVal 161
             |||:::||||||||||||:::|||:::      |||||||||   |||||||||||||||
Db       602 CCACTTTACATCTCTGAAATGGCTCATGAGCGTGTCAGAGGAACTTTGGGATCCTGTGTG 661

Qy       162 GlnLeuMetValValValGlyIleLeuLeuAlaTyrLeuAlaGlyTrpValLeuGluTrp 181
             |||||||||||||||:::||||||:::   ||||||:::   |||      |||:::|||
Db       662 CAACTAATGGTTGTGATAGGCATTATGGGTGCTTATGTAACAGGTTTGTTCCTGGACTGG 721

Qy       182 ArgTrpLeuAlaValLeuGlyCysValProProSerLeuMetLeuLeuLeuMetCysPhe 201
             |||||||||||||||         :::||||||:::||||||||||||   |||||||||
Db       722 CGATGGTTAGCAGTGGCATCATCTATTCCTCCCACACTCATGCTGCTGTCCATGTGCTTC 781

Qy       202 MetProGluThrProArgPheLeuLeuThrGlnHisArgArgGlnGluAlaMetAlaAla 221
             |||||||||||||||||||||||||||   |||   :::|||:::||||||      |||
Db       782 ATGCCGGAGACGCCACGATTTCTGCTCTGCCAAGGGAAACGGCGGGAAGCAGAAGACGCC 841

Qy       222 LeuArgPheLeuTrpGlySerGlu------GlnGlyTrp---------------GluAsp 234
             ||||||||||||   |||   :::         |||                  ||||||
Db       842 CTGCGCTTCCTTAGGGGTCCTGATGCCCCCGCCGAGTGGGAATGCGCTAGAATAGAAGAT 901

Qy       235 ProProIleGlyAlaGluGlnSerPheHisLeuAlaLeuLeuArgGlnProGlyIleTyr 254
                            ||||||||||||   |||      |||:::   ||||||:::|||
Db       902 GCCTACAAAAATGAAGAGCAAAGCTTCAGTCTGGGGGATCTGAAGGACCCTGGTGTCTAT 961

Qy       255 LysProPheIleIleGlyValSerLeuMetAlaPheGlnGlnLeuSerGlyValAsnAla 274
             ||||||      |||||||||   :::|||      ||||||   :::|||:::||||||
Db       962 AAACCTTTGGGCATTGGTGTTATGATGATGCTCCTCCAGCAGTTTACCGGCATCAACGCC 1021

Qy       275 ValMetPheTyrAlaGluThrIlePheGluGluAlaLysPheLysAspSerSerLeuAla 294
             :::|||||||||||||||||||||||||||:::|||   ||||||   |||   :::|||
Db      1022 ATCATGTTTTACGCAGAGACCATCTTCGAGCAGGCCCATTTCAAAAGCAGTGATGTGGCC 1081

Qy       295 SerValValValGlyValIleGlnValLeuPheThrAlaValAlaAlaLeuIleMetAsp 314
             :::|||:::|||         ||||||:::|||||||||:::||||||||||||||||||
Db      1082 ACTGTTATTGTCGCAGCCACACAGGTGGTCTTCACTGCAATTGCTGCTCTTATAATGGAC 1141

Qy       315 ArgAlaGlyArgArgLeuLeuLeuValLeuSerGlyValValMetValPheSerThrSer 334
             :::|||||||||::::::||||||:::||||||||||||||||||      |||   :::
Db      1142 AAGGCCGGGCGGAAAGTTCTTCTCATTTTGTCTGGTGTCGTCATGTGTGTTAGCGAAGCA 1201

Qy       335 AlaPheGlyAlaTyrPheLysLeuThrGlnGlyGlyProGlyAsnSerSerHisValAla 354
                ||||||   |||||||||||||||         |||   |||||||||      :::
Db      1202 GTGTTTGGAGTGTACTTTAAGCTGACTGTGATGAAGCCCAATAACTCCTCAATGACCAGT 1261

Qy       355 IleSerAlaPro------------ValSerAlaGlnProValAspAlaSerValGlyLeu 370
             :::                     :::      ||||||         |||      |||
Db      1262 GTGCTAACGGACACCCATGGTCTTCTGGAAGATCAGCCC---------TCTGCAGACCTG 1312

Qy       371 AlaTrpLeuAlaValGlySerMetCysLeuPheIleAlaGlyPheAlaValGlyTrpGly 390
             ||||||||||||||||||||||||      ||||||||||||||||||:::|||||||||
Db      1313 GCCTGGTTGGCTGTGGGAAGCATGGGCTTTTTCATTGCAGGATTTGCTATTGGTTGGGGT 1372

Qy       391 ProIleProTrpLeuLeuMetSerGluIlePheProLeuHisValLysGlyValAlaThr 410
             |||   |||||||||:::||||||||||||||||||      |||:::|||:::   :::
Db      1373 CCTACTCCGTGGCTGGTGATGTCCGAGATCTTCCCCACGCGGGTGAGGGGATTAGGCAGT 1432

Qy       411 GlyIleCysValLeuThrAsnTrpLeuMetAlaPheLeuValThrLysGluPheSerSer 430
                :::||||||||||||||||||      ||||||:::|||||||||   |||   :::
Db      1433 GCTCTGTGTGTGCTTACTAACTGGACCTGTGCCTTCATTGTCACCAAAACCTTCCAGAAC 1492

Qy       431 LeuMetGluValLeuArgProTyrGlyAlaPheTrpLeuAlaSerAlaPheCysIlePhe 450
             ||||||:::   |||         |||   ||||||:::   ||||||   |||
Db      1493 CTCATGGATGCTCTGAGCAGCGCAGGAACATTTTGGATGTTCTCGGCACTGTGTGCTTCT 1552

Qy       451 SerValLeuPheThrLeuPheCysValProGluThrLysGlyLysThrLeuGluGlnIle 470
             :::|||:::||||||   |||   ||||||||||||||||||||||||||||||:::|||
Db      1553 AATGTCGTCTTCACGGCTTTCTTTGTCCCTGAGACCAAAGGCAAAACTCTGGAAGAGATT 1612

Qy       471 ThrAlaHisPheGluGly 476
                |||   |||:::|||
Db      1613 CAAGCTGGTTTTAAAGGG 1630
```

```
Search completed: September 28, 2004, 03:28:36
Job time : 4617 secs
```

```
                    GenCore version 5.1.6
             Copyright (c) 1993 - 2004  Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on:         September 28, 2004, 03:28:39 ; Search time 4236 Seconds
                                            (without alignments)
                                            570.345 Million cell updates/sec

Title:          US-09-886-954A-1
Perfect score:  2457
Sequence:       1 MTPEDPEETQPLLGPPGGSA..........CVPETKGKTLEQITAHFEGR 477

Scoring table:  BLOSUM62
                Xgapop 10.0 , Xgapext  0.5
                Ygapop 10.0 , Ygapext  0.5
                Fgapop  6.0 , Fgapext  7.0
                Delop   6.0 , Delext   7.0

Searched:       3337386 seqs, 2532474682 residues

Total number of hits satisfying chosen parameters:      6674772

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US09886954/runat_27092004_170634_13496/app_query.fasta_1.647
-DB=Published_Applications_NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09886954_@CGN_1_1_783_@runat_27092004_170634_13496
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :      Published_Applications_NA:*
                1:  /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
                2:  /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
                3:  /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
                4:  /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
                5:  /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
                6:  /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
                7:  /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
                8:  /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
                9:  /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
                10:  /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
                11:  /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
                12:  /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
                13:  /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq2:*
                14:  /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
                15:  /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
                16:  /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
                17:  /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
                18:  /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
                19:  /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                              SUMMARIES
                   %
Result             Query
   No.    Score    Match Length DB  ID                       Description
----------------------------------------------------------------------------------
   1      2457   100.0    1862  16  US-10-115-831-137          Sequence 137, App
   2      2457   100.0    2080  15  US-10-168-651-28           Sequence 28, Appl
   3    2067.5    84.1    1655  16  US-10-115-831-138          Sequence 138, App
   4    1679.5    68.4    1156  16  US-10-264-237-688          Sequence 688, App
   5    1554.5    63.3    1461  13  US-10-169-395-29           Sequence 29, Appl
   6      1268    51.6     789  13  US-10-169-395-19           Sequence 19, Appl
   7     958.5    39.0    2356  15  US-10-157-031-358          Sequence 358, App
   8     958.5    39.0    2487  13  US-10-342-887-1685         Sequence 1685, Ap
   9     958.5    39.0    2487  13  US-10-172-118-1685         Sequence 1685, Ap
  10     958.5    39.0    2487  17  US-10-755-889-109          Sequence 109, App
  11     958.5    39.0    2696  10  US-09-814-353-19608        Sequence 19608, A
  12       716    29.1    2202  13  US-10-424-599-132422       Sequence 132422,
  13       702    28.6    1797  13  US-10-425-114-26596        Sequence 26596, A
  14       702    28.6    2003  17  US-10-437-963-61061        Sequence 61061, A
  15     701.5    28.6    2049  13  US-10-425-114-24968        Sequence 24968, A
  16       682    27.8    1875  13  US-10-425-114-26550        Sequence 26550, A
  17     666.5    27.1    1953  14  US-10-051-909-35           Sequence 35, Appl
  18       639    26.0    1768  16  US-10-310-154-357          Sequence 357, App
  19     620.5    25.3  150130  13  US-10-087-192-820          Sequence 820, App
  20     609.5    24.8    1785  17  US-10-437-963-64291        Sequence 64291, A
  21       607    24.7    1386  16  US-10-369-493-47058        Sequence 47058, A
  22     604.5    24.6    2045  13  US-10-425-114-22342        Sequence 22342, A
  23     604.5    24.6    2051  13  US-10-425-114-29817        Sequence 29817, A
  24     604.5    24.6    2190  17  US-10-437-963-74980        Sequence 74980, A
  25       603    24.5    1419  16  US-10-369-493-47223        Sequence 47223, A
  26     598.5    24.4    1395  13  US-10-282-122A-20182       Sequence 20182, A
  27     593.5    24.2    1766  13  US-10-425-114-22366        Sequence 22366, A
  28     593.5    24.2    1767  17  US-10-437-963-22036        Sequence 22036, A
  29     591.5    24.1    1395  13  US-10-282-122A-6734        Sequence 6734, Ap
  30     591.5    24.1    1395  16  US-10-369-493-47239        Sequence 47239, A
  31     587.5    23.9    1359  9  US-09-938-842A-1627       Sequence 1627, Ap
  32     587.5    23.9    1359  11  US-09-938-842A-1627        Sequence 1627, Ap
  33     584.5    23.8    1473  9  US-09-938-842A-2331       Sequence 2331, Ap
  34     584.5    23.8    1473  11  US-09-938-842A-2331        Sequence 2331, Ap
  35     583.5    23.7    1395  13  US-10-282-122A-38732       Sequence 38732, A
  36     583.5    23.7    1395  13  US-10-282-122A-40124       Sequence 40124, A
  37       582    23.7    1566  9  US-09-938-842A-1769       Sequence 1769, Ap
  38       582    23.7    1566  11  US-09-938-842A-1769        Sequence 1769, Ap
  39       582    23.7    1898  13  US-10-424-599-96711        Sequence 96711, A
  40     574.5    23.4    1673  17  US-10-664-705-79           Sequence 79, Appl
  41       571    23.2    1338  13  US-10-282-122A-35021       Sequence 35021, A
  42     570.5    23.2    3915  9  US-09-954-456-507        Sequence 507, App
  43     570.5    23.2    3915  13  US-10-342-887-1341         Sequence 1341, Ap
  44     570.5    23.2    3915  13  US-10-172-118-1341         Sequence 1341, Ap
  45     570.5    23.2    3915  13  US-10-170-385-248          Sequence 248, App


                              ALIGNMENTS

RESULT 1
US-10-115-831-137
; Sequence 137, Application US/10115831
; Publication No. US20030219743A1
; GENERAL INFORMATION:
;  APPLICANT: Tang, Y. Tom
;  APPLICANT:  Liu, Chenghua
;  APPLICANT:  Asundi, Vinod
;  APPLICANT:  Ren, Feiyan
;  APPLICANT:  Drmanac, Radoje T.
;  TITLE OF INVENTION: No. US20030219743A1el Nucleic Acids and
;  TITLE OF INVENTION:  Polypeptides
;  FILE REFERENCE: 792CIP2ADIV
;  CURRENT APPLICATION NUMBER: US/10/115,831
;  CURRENT FILING DATE:  2002-04-02
;  PRIOR APPLICATION NUMBER: 09/667,298
;  PRIOR FILING DATE: 2000-09-22
;  PRIOR APPLICATION NUMBER: 09/577,408
;  PRIOR FILING DATE: 2000-05-18
;  NUMBER OF SEQ ID NOS: 178
;  SOFTWARE: pt_FL_genes Version 2.0
; SEQ ID NO 137
;   LENGTH: 1862
```

```
;    TYPE: DNA
;    ORGANISM: Homo sapiens
;    FEATURE:
;    NAME/KEY: CDS
;    LOCATION: (4)..(1437)
US-10-115-831-137

Alignment Scores:
Pred. No.:              3e-254        Length:          1862
Score:                  2457.00       Matches:         477
Percent Similarity:     100.00%       Conservative:    0
Best Local Similarity:  100.00%       Mismatches:      0
Query Match:            100.00%       Indels:          0
DB:                     16            Gaps:            0

US-09-886-954A-1 (1-477) x US-10-115-831-137 (1-1862)

Qy          1 MetThrProGluAspProGluGluThrGlnProLeuLeuGlyProProGlyGlySerAla 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          4 ATGACGCCCGAGGACCCAGAGGAAACCCAGCCGCTTCTGGGGCCTCCTGGCGGCAGCGCG 63

Qy         21 ProArgGlyArgArgValPheLeuAlaAlaPheAlaAlaAlaLeuGlyProLeuSerPhe 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         64 CCCCGCGGCCGCCGCGTCTTCCTCGCCGCCTTCGCCGCTGCCCTGGGCCCACTCAGCTTC 123

Qy         41 GlyPheAlaLeuGlyTyrSerSerProAlaIleProSerLeuGlnArgAlaAlaProPro 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        124 GGCTTCGCGCTCGGCTACAGCTCCCCGGCCATCCCTAGCCTGCAGCGCGCCGCGCCCCCG 183

Qy         61 AlaProArgLeuAspAspAlaAlaAlaSerTrpPheGlyAlaValValThrLeuGlyAla 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        184 GCCCCGCGCCTGGACGACGCCGCCGCCTCCTGGTTCGGGGCTGTCGTGACCCTGGGTGCC 243

Qy         81 AlaAlaGlyGlyValLeuGlyGlyTrpLeuValAspArgAlaGlyArgLysLeuSerLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        244 GCGGCGGGGGGAGTGCTGGGCGGCTGGCTGGTGGACCGCGCCGGGCGCAAGCTGAGCCTC 303

Qy        101 LeuLeuCysSerValProPheValAlaGlyPheAlaValIleThrAlaAlaGlnAspVal 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        304 TTGCTGTGCTCCGTGCCCTTCGTGGCCGGCTTTGCCGTCATCACCGCGGCCCAGGACGTG 363

Qy        121 TrpMetLeuLeuGlyGlyArgLeuLeuThrGlyLeuAlaCysGlyValAlaSerLeuVal 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        364 TGGATGCTGCTGGGGGGCCGCCTCCTCACCGGCCTGGCCTGCGGTGTTGCCTCCCTAGTG 423

Qy        141 AlaProValTyrIleSerGluIleAlaTyrProAlaValArgGlyLeuLeuGlySerCys 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        424 GCCCCGGTCTACATCTCCGAAATCGCCTACCCAGCAGTCCGGGGGTTGCTCGGCTCCTGT 483

Qy        161 ValGlnLeuMetValValValGlyIleLeuLeuAlaTyrLeuAlaGlyTrpValLeuGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        484 GTGCAGCTAATGGTCGTCGTCGGCATCCTCCTGGCCTACCTGGCAGGCTGGGTGCTGGAG 543

Qy        181 TrpArgTrpLeuAlaValLeuGlyCysValProProSerLeuMetLeuLeuLeuMetCys 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        544 TGGCGCTGGCTGGCTGTGCTGGGCTGCGTGCCCCCCTCCCTCATGCTGCTTCTCATGTGC 603

Qy        201 PheMetProGluThrProArgPheLeuLeuThrGlnHisArgArgGlnGluAlaMetAla 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        604 TTCATGCCCGAGACCCCGCGCTTCCTGCTGACTCAGCACAGGCGCCAGGAGGCCATGGCC 663

Qy        221 AlaLeuArgPheLeuTrpGlySerGluGlnGlyTrpGluAspProProIleGlyAlaGlu 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        664 GCCCTGCGGTTCCTGTGGGGCTCCGAGCAGGGCTGGGAAGACCCCCCCATCGGGGCTGAG 723

Qy        241 GlnSerPheHisLeuAlaLeuLeuArgGlnProGlyIleTyrLysProPheIleIleGly 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        724 CAGAGCTTTCACCTGGCCCTGCTGCGGCAGCCCGGCATCTACAAGCCCTTCATCATCGGT 783

Qy        261 ValSerLeuMetAlaPheGlnGlnLeuSerGlyValAsnAlaValMetPheTyrAlaGlu 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        784 GTCTCCCTGATGGCCTTCCAGCAGCTGTCGGGGGTCAACGCCGTCATGTTCTATGCAGAG 843

Qy        281 ThrIlePheGluGluAlaLysPheLysAspSerSerLeuAlaSerValValValGlyVal 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        844 ACCATCTTTGAAGAGGCCAAGTTCAAGGACAGCAGCCTGGCCTCGGTCGTCGTGGGTGTC 903

Qy        301 IleGlnValLeuPheThrAlaValAlaAlaLeuIleMetAspArgAlaGlyArgArgLeu 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        904 ATCCAGGTGCTGTTCACAGCTGTGGCGGCTCTCATCATGGACAGAGCAGGGCGGAGGCTG 963

Qy        321 LeuLeuValLeuSerGlyValValMetValPheSerThrSerAlaPheGlyAlaTyrPhe 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        964 CTCCTGGTCTTGTCAGGTGTGGTCATGGTGTTCAGCACGAGTGCCTTCGGCGCCTACTTC 1023

Qy        341 LysLeuThrGlnGlyGlyProGlyAsnSerSerHisValAlaIleSerAlaProValSer 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1024 AAGCTGACCCAGGGTGGCCCTGGCAACTCCTCGCACGTGGCCATCTCGGCGCCTGTCTCT 1083

Qy        361 AlaGlnProValAspAlaSerValGlyLeuAlaTrpLeuAlaValGlySerMetCysLeu 380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1084 GCACAGCCTGTTGATGCCAGCGTGGGGCTGGCCTGGCTGGCCGTGGGCAGCATGTGCCTC 1143

Qy        381 PheIleAlaGlyPheAlaValGlyTrpGlyProIleProTrpLeuLeuMetSerGluIle 400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1144 TTCATCGCCGGCTTTGCGGTGGGCTGGGGGCCCATCCCCTGGCTCCTCATGTCAGAGATC 1203

Qy        401 PheProLeuHisValLysGlyValAlaThrGlyIleCysValLeuThrAsnTrpLeuMet 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1204 TTCCCTCTGCATGTCAAGGGCGTGGCGACAGGCATCTGCGTCCTCACCAACTGGCTCATG 1263

Qy        421 AlaPheLeuValThrLysGluPheSerSerLeuMetGluValLeuArgProTyrGlyAla 440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1264 GCCTTTCTCGTGACCAAGGAGTTCAGCAGCCTCATGGAGGTCCTCAGGCCCTATGGAGCC 1323

Qy        441 PheTrpLeuAlaSerAlaPheCysIlePheSerValLeuPheThrLeuPheCysValPro 460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1324 TTCTGGCTTGCCTCCGCTTTCTGCATCTTCAGTGTCCTTTTCACTTTGTTCTGTGTCCCT 1383

Qy        461 GluThrLysGlyLysThrLeuGluGlnIleThrAlaHisPheGluGlyArg 477
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1384 GAAACTAAAGGAAAGACTCTGGAACAAATCACAGCCCATTTTGAGGGGCGA 1434

RESULT 2
US-10-168-651-28
; Sequence 28, Application US/10168651
; Publication No. US20030171275A1
; GENERAL INFORMATION:
;  APPLICANT: INCYTE GENOMICS, INC.
;  APPLICANT:  BAUGHN, Mariah R.
;  APPLICANT:  BURFORD, Neil
;  APPLICANT:  AU-YOUNG, Janice
;  APPLICANT:  LU, Dyung Aina M.
;  APPLICANT:  YANG, Junming
;  APPLICANT:  REDDY, Roopa
;  APPLICANT:  LAL, Preeti
;  APPLICANT:  HILLMAN, Jennifer L.
;  APPLICANT:  AZIMZAI, Yalda
;  APPLICANT:  YUE, Henry
;  APPLICANT:  NGUYEN, Danniel B.
;  APPLICANT:  YAO, Monique G.
;  APPLICANT:  GANDHI, Ameeni R.
;  APPLICANT:  TANG, Y. Tom
;  APPLICANT:  KHAN, Farrah A.
;  TITLE OF INVENTION: TRANSPORTERS AND ION CHANNELS
;  FILE REFERENCE: PI-0005 PCT
;  CURRENT APPLICATION NUMBER: US/10/168,651
;  CURRENT FILING DATE:  2002-06-21
;  PRIOR APPLICATION NUMBER: 60/172,000; 60/176,083; 60/177,332; 60/178,572; 60/179,758
;    60/181,625
;  PRIOR FILING DATE: 1999-12-23; 2000-01-14; 2000-01-21; 2000-01-28; 2000-02-02;
;    2000-02-10
;  NUMBER OF SEQ ID NOS: 54
;  SOFTWARE: PERL Program
; SEQ ID NO 28
;   LENGTH: 2080
```

```
;    TYPE: DNA
;    ORGANISM: Homo sapiens
;    FEATURE:
;    NAME/KEY: misc_feature
;    OTHER INFORMATION: Incyte ID No. US20030171275A1 1416107CB1
US-10-168-651-28

Alignment Scores:
Pred. No.:                 3.51e-254      Length:        2080
Score:                     2457.00        Matches:       477
Percent Similarity:        100.00%        Conservative:  0
Best Local Similarity:     100.00%        Mismatches:    0
Query Match:               100.00%        Indels:        0
DB:                        15             Gaps:          0

US-09-886-954A-1 (1-477) x US-10-168-651-28 (1-2080)

Qy          1 MetThrProGluAspProGluGluThrGlnProLeuLeuGlyProProGlyGlySerAla 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         46 ATGACGCCCGAGGACCCAGAGGAAACCCAGCCGCTTCTGGGGCCTCCTGGCGGCAGCGCG 105

Qy         21 ProArgGlyArgArgValPheLeuAlaAlaPheAlaAlaAlaLeuGlyProLeuSerPhe 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        106 CCCCGCGGCCGCCGCGTCTTCCTCGCCGCCTTCGCCGCTGCCCTGGGCCCACTCAGCTTC 165

Qy         41 GlyPheAlaLeuGlyTyrSerSerProAlaIleProSerLeuGlnArgAlaAlaProPro 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        166 GGCTTCGCGCTCGGCTACAGCTCCCCGGCCATCCCTAGCCTGCAGCGCGCCGCGCCCCCG 225

Qy         61 AlaProArgLeuAspAspAlaAlaAlaSerTrpPheGlyAlaValValThrLeuGlyAla 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        226 GCCCCGCGCCTGGACGACGCCGCCGCCTCCTGGTTCGGGGCTGTCGTGACCCTGGGTGCC 285

Qy         81 AlaAlaGlyGlyValLeuGlyGlyTrpLeuValAspArgAlaGlyArgLysLeuSerLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        286 GCGGCGGGGGGAGTGCTGGGCGGCTGGCTGGTGGACCGCGCCGGGCGCAAGCTGAGCCTC 345

Qy        101 LeuLeuCysSerValProPheValAlaGlyPheAlaValIleThrAlaAlaGlnAspVal 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        346 TTGCTGTGCTCCGTGCCCTTCGTGGCCGGCTTTGCCGTCATCACCGCGGCCCAGGACGTG 405

Qy        121 TrpMetLeuLeuGlyGlyArgLeuLeuThrGlyLeuAlaCysGlyValAlaSerLeuVal 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        406 TGGATGCTGCTGGGGGGCCGCCTCCTCACCGGCCTGGCCTGCGGTGTTGCCTCCCTAGTG 465

Qy        141 AlaProValTyrIleSerGluIleAlaTyrProAlaValArgGlyLeuLeuGlySerCys 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        466 GCCCCGGTCTACATCTCCGAAATCGCCTACCCAGCAGTCCGGGGGTTGCTCGGCTCCTGT 525

Qy        161 ValGlnLeuMetValValValGlyIleLeuLeuAlaTyrLeuAlaGlyTrpValLeuGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        526 GTGCAGCTAATGGTCGTCGTCGGCATCCTCCTGGCCTACCTGGCAGGCTGGGTGCTGGAG 585

Qy        181 TrpArgTrpLeuAlaValLeuGlyCysValProProSerLeuMetLeuLeuLeuMetCys 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        586 TGGCGCTGGCTGGCTGTGCTGGGCTGCGTGCCCCCCTCCCTCATGCTGCTTCTCATGTGC 645

Qy        201 PheMetProGluThrProArgPheLeuLeuThrGlnHisArgArgGlnGluAlaMetAla 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        646 TTCATGCCCGAGACCCCGCGCTTCCTGCTGACTCAGCACAGGCGCCAGGAGGCCATGGCC 705

Qy        221 AlaLeuArgPheLeuTrpGlySerGluGlnGlyTrpGluAspProProIleGlyAlaGlu 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        706 GCCCTGCGGTTCCTGTGGGGCTCCGAGCAGGGCTGGGAAGACCCCCCCATCGGGGCTGAG 765

Qy        241 GlnSerPheHisLeuAlaLeuLeuArgGlnProGlyIleTyrLysProPheIleIleGly 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        766 CAGAGCTTTCACCTGGCCCTGCTGCGGCAGCCCGGCATCTACAAGCCCTTCATCATCGGC 825

Qy        261 ValSerLeuMetAlaPheGlnGlnLeuSerGlyValAsnAlaValMetPheTyrAlaGlu 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        826 GTCTCCCTGATGGCCTTCCAGCAGCTGTCGGGGGTCAACGCCGTCATGTTCTATGCAGAG 885

Qy        281 ThrIlePheGluGluAlaLysPheLysAspSerSerLeuAlaSerValValValGlyVal 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        886 ACCATCTTTGAAGAGGCCAAGTTCAAGGACAGCAGCCTGGCCTCGGTCGTCGTGGGTGTC 945

Qy        301 IleGlnValLeuPheThrAlaValAlaAlaLeuIleMetAspArgAlaGlyArgArgLeu 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        946 ATCCAGGTGCTGTTCACAGCTGTGGCGGCTCTCATCATGGACAGAGCAGGGCGGAGGCTG 1005

Qy        321 LeuLeuValLeuSerGlyValValMetValPheSerThrSerAlaPheGlyAlaTyrPhe 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1006 CTCCTGGTCTTGTCAGGTGTGGTCATGGTGTTCAGCACGAGTGCCTTCGGCGCCTACTTC 1065

Qy        341 LysLeuThrGlnGlyGlyProGlyAsnSerSerHisValAlaIleSerAlaProValSer 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1066 AAGCTGACCCAGGGTGGCCCTGGCAACTCCTCGCACGTGGCCATCTCGGCGCCTGTCTCT 1125

Qy        361 AlaGlnProValAspAlaSerValGlyLeuAlaTrpLeuAlaValGlySerMetCysLeu 380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1126 GCACAGCCTGTTGATGCCAGCGTGGGGCTGGCCTGGCTGGCCGTGGGCAGCATGTGCCTC 1185

Qy        381 PheIleAlaGlyPheAlaValGlyTrpGlyProIleProTrpLeuLeuMetSerGluIle 400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1186 TTCATCGCCGGCTTTGCGGTGGGCTGGGGGCCCATCCCCTGGCTCCTCATGTCAGAGATC 1245

Qy        401 PheProLeuHisValLysGlyValAlaThrGlyIleCysValLeuThrAsnTrpLeuMet 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1246 TTCCCTCTGCATGTCAAGGGCGTGGCGACAGGCATCTGCGTCCTCACCAACTGGCTCATG 1305

Qy        421 AlaPheLeuValThrLysGluPheSerSerLeuMetGluValLeuArgProTyrGlyAla 440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1306 GCCTTTCTCGTGACCAAGGAGTTCAGCAGCCTCATGGAGGTCCTCAGGCCCTATGGAGCC 1365

Qy        441 PheTrpLeuAlaSerAlaPheCysIlePheSerValLeuPheThrLeuPheCysValPro 460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1366 TTCTGGCTTGCCTCCGCTTTCTGCATCTTCAGTGTCCTTTTCACTTTGTTCTGTGTCCCT 1425

Qy        461 GluThrLysGlyLysThrLeuGluGlnIleThrAlaHisPheGluGlyArg 477
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1426 GAAACTAAAGGAAAGACTCTGGAACAAATCACAGCCCATTTTGAGGGGCGA 1476

RESULT 3
US-10-115-831-138
; Sequence 138, Application US/10115831
; Publication No. US20030219743A1
; GENERAL INFORMATION:
;  APPLICANT: Tang, Y. Tom
;  APPLICANT:  Liu, Chenghua
;  APPLICANT:  Asundi, Vinod
;  APPLICANT:  Ren, Feiyan
;  APPLICANT:  Drmanac, Radoje T.
;  TITLE OF INVENTION: No. US20030219743Alel Nucleic Acids and
;  TITLE OF INVENTION:  Polypeptides
;  FILE REFERENCE: 792CIP2ADIV
;  CURRENT APPLICATION NUMBER: US/10/115,831
;  CURRENT FILING DATE:  2002-04-02
;  PRIOR APPLICATION NUMBER: 09/667,298
;  PRIOR FILING DATE: 2000-09-22
;  PRIOR APPLICATION NUMBER: 09/577,408
;  PRIOR FILING DATE: 2000-05-18
;  NUMBER OF SEQ ID NOS: 178
;  SOFTWARE: pt_FL_genes Version 2.0
; SEQ ID NO 138
;   LENGTH: 1655
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (4)..(1230)
US-10-115-831-138

Alignment Scores:
Pred. No.:                 2.25e-212      Length:        1655
Score:                     2067.50        Matches:       408
```

```
Percent Similarity:      85.53%        Conservative:   0
Best Local Similarity:   85.53%        Mismatches:     0
Query Match:             84.15%        Indels:         69
DB:                      16            Gaps:           1

US-09-886-954A-1 (1-477) x US-10-115-831-138 (1-1655)

Qy          1 MetThrProGluAspProGluGluThrGlnProLeuLeuGlyProProGlyGlySerAla 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          4 ATGACGCCCGAGGACCCAGAGGAAACCCAGCCGCTTCTGGGGCCTCCTGGCGGCAGCGCG 63

Qy         21 ProArgGlyArgArgValPheLeuAlaAlaPheAlaAlaAlaLeuGlyProLeuSerPhe 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         64 CCCCGCGGCCGCCGCGTCTTCCTCGCCGCCTTCGCCGCTGCCCTGGGCCCACTCAGCTTC 123

Qy         41 GlyPheAlaLeuGlyTyrSerSerProAlaIleProSerLeuGlnArgAlaAlaProPro 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        124 GGCTTCGCGCTCGGCTACAGCTCCCCGGCCATCCCTAGCCTGCAGCGCGCCGCGCCCCCG 183

Qy         61 AlaProArgLeuAspAspAlaAlaAlaSerTrpPheGlyAlaValValThrLeuGlyAla 80
              |||||||||||||||||||||||||||||||||||||||
Db        184 GCCCCGCGCCTGGACGACGCCGCCGCCTCCTGGTTCGGG--------------------- 222

Qy         81 AlaAlaGlyGlyValLeuGlyGlyTrpLeuValAspArgAlaGlyArgLysLeuSerLeu 100

Db        222 ------------------------------------------------------------ 222

Qy        101 LeuLeuCysSerValProPheValAlaGlyPheAlaValIleThrAlaAlaGlnAspVal 120

Db        222 ------------------------------------------------------------ 222

Qy        121 TrpMetLeuLeuGlyGlyArgLeuLeuThrGlyLeuAlaCysGlyValAlaSerLeuVal 140

Db        222 ------------------------------------------------------------ 222

Qy        141 AlaProValTyrIleSerGluIleAlaTyrProAlaValArgGlyLeuLeuGlySerCys 160
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        223 ------GTCTACATCTCCGAAATCGCCTACCCAGCAGTCCGGGGGTTGCTCGGCTCCTGT 276

Qy        161 ValGlnLeuMetValValValGlyIleLeuLeuAlaTyrLeuAlaGlyTrpValLeuGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        277 GTGCAGCTAATGGTCGTCGTCGGCATCCTCCTGGCCTACCTGGCAGGCTGGGTGCTGGAG 336

Qy        181 TrpArgTrpLeuAlaValLeuGlyCysValProProSerLeuMetLeuLeuLeuMetCys 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        337 TGGCGCTGGCTGGCTGTGCTGGGCTGCGTGCCCCCCTCCCTCATGCTGCTTCTCATGTGC 396

Qy        201 PheMetProGluThrProArgPheLeuLeuThrGlnHisArgArgGlnGluAlaMetAla 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        397 TTCATGCCCGAGACCCCGCGCTTCCTGCTGACTCAGCACAGGCGCCAGGAGGCCATGGCC 456

Qy        221 AlaLeuArgPheLeuTrpGlySerGluGlnGlyTrpGluAspProProIleGlyAlaGlu 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        457 GCCCTGCGGTTCCTGTGGGGCTCCGAGCAGGGCTGGGAAGACCCCCCCATCGGGGCTGAG 516

Qy        241 GlnSerPheHisLeuAlaLeuLeuArgGlnProGlyIleTyrLysProPheIleIleGly 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        517 CAGAGCTTTCACCTGGCCCTGCTGCGGCAGCCCGGCATCTACAAGCCCTTCATCATCGGT 576

Qy        261 ValSerLeuMetAlaPheGlnGlnLeuSerGlyValAsnAlaValMetPheTyrAlaGlu 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        577 GTCTCCCTGATGGCCTTCCAGCAGCTGTCGGGGGTCAACGCCGTCATGTTCTATGCAGAG 636

Qy        281 ThrIlePheGluGluAlaLysPheLysAspSerSerLeuAlaSerValValValGlyVal 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        637 ACCATCTTTGAAGAGGCCAAGTTCAAGGACAGCAGCCTGGCCTCGGTCGTCGTGGGTGTC 696

Qy        301 IleGlnValLeuPheThrAlaValAlaAlaLeuIleMetAspArgAlaGlyArgArgLeu 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        697 ATCCAGGTGCTGTTCACAGCTGTGGCGGCTCTCATCATGGACAGAGCAGGGCGGAGGCTG 756

Qy        321 LeuLeuValLeuSerGlyValValMetValPheSerThrSerAlaPheGlyAlaTyrPhe 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        757 CTCCTGGTCTTGTCAGGTGTGGTCATGGTGTTCAGCACGAGTGCCTTCGGCGCCTACTTC 816

Qy        341 LysLeuThrGlnGlyGlyProGlyAsnSerSerHisValAlaIleSerAlaProValSer 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        817 AAGCTGACCCAGGGTGGCCCTGGCAACTCCTCGCACGTGGCCATCTCGGCGCCTGTCTCT 876

Qy        361 AlaGlnProValAspAlaSerValGlyLeuAlaTrpLeuAlaValGlySerMetCysLeu 380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        877 GCACAGCCTGTTGATGCCAGCGTGGGGCTGGCCTGGCTGGCCGTGGGCAGCATGTGCCTC 936

Qy        381 PheIleAlaGlyPheAlaValGlyTrpGlyProIleProTrpLeuLeuMetSerGluIle 400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        937 TTCATCGCCGGCTTTGCGGTGGGCTGGGGGCCCATCCCCTGGCTCCTCATGTCAGAGATC 996

Qy        401 PheProLeuHisValLysGlyValAlaThrGlyIleCysValLeuThrAsnTrpLeuMet 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        997 TTCCCTCTGCATGTCAAGGGCGTGGCGACAGGCATCTGCGTCCTCACCAACTGGCTCATG 1056

Qy        421 AlaPheLeuValThrLysGluPheSerSerLeuMetGluValLeuArgProTyrGlyAla 440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1057 GCCTTTCTCGTGACCAAGGAGTTCAGCAGCCTCATGGAGGTCCTCAGGCCCTATGGAGCC 1116

Qy        441 PheTrpLeuAlaSerAlaPheCysIlePheSerValLeuPheThrLeuPheCysValPro 460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1117 TTCTGGCTTGCCTCCGCTTTCTGCATCTTCAGTGTCCTTTTCACTTTGTTCTGTGTCCCT 1176

Qy        461 GluThrLysGlyLysThrLeuGluGlnIleThrAlaHisPheGluGlyArg 477
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1177 GAAACTAAAGGAAAGACTCTGGAACAAATCACAGCCCATTTTGAGGGGCGA 1227

RESULT 4
US-10-264-237-688
; Sequence 688, Application US/10264237
; Publication No. US20040009491A1
; GENERAL INFORMATION:
;  APPLICANT: Birse et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PA131P1
;  CURRENT APPLICATION NUMBER: US/10/264,237
;  CURRENT FILING DATE:  2002-10-04
;  PRIOR APPLICATION NUMBER: PCT/US01/16450
;  PRIOR FILING DATE: 2001-05-18
;  PRIOR APPLICATION NUMBER: US 60/205,515
;  PRIOR FILING DATE: 2000-05-19
;  NUMBER OF SEQ ID NOS: 2876
;  SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 688
;   LENGTH: 1156
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (24)..(24)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (47)..(47)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (351)..(351)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (397)..(397)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (432)..(432)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (792)..(792)
```

```
;    OTHER INFORMATION: n equals a,t,g, or c
;    FEATURE:
;    NAME/KEY: misc_feature
;    LOCATION: (1119)..(1119)
;    OTHER INFORMATION: n equals a,t,g, or c
;    FEATURE:
;    NAME/KEY: misc_feature
;    LOCATION: (1156)..(1156)
;    OTHER INFORMATION: n equals a,t,g, or c
US-10-264-237-688

Alignment Scores:
Pred. No.:                8.28e-171       Length:          1156
Score:                    1679.50         Matches:         344
Percent Similarity:       72.12%          Conservative:    0
Best Local Similarity:    72.12%          Mismatches:      10
Query Match:              68.36%          Indels:          125
DB:                       16              Gaps:            1

US-09-886-954A-1 (1-477) x US-10-264-237-688 (1-1156)

Qy          1 MetThrProGluAspProGluGluThrGlnProLeuLeuGlyProProGlyGlySerAla 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         53 ATGACGCCCGAGGACCCAGAGGAAACCCAGCCGCTTCTGGGGCCTCCTGGCGGCAR---- 108

Qy         21 ProArgGlyArgArgValPheLeuAlaAlaPheAlaAlaAlaLeuGlyProLeuSerPhe 40

Db        108 ------------------------------------------------------------ 108

Qy         41 GlyPheAlaLeuGlyTyrSerSerProAlaIleProSerLeuGlnArgAlaAlaProPro 60

Db        108 ------------------------------------------------------------ 108

Qy         61 AlaProArgLeuAspAspAlaAlaAlaSerTrpPheGlyAlaValValThrLeuGlyAla 80

Db        108 ------------------------------------------------------------ 108

Qy         81 AlaAlaGlyGlyValLeuGlyGlyTrpLeuValAspArgAlaGlyArgLysLeuSerLeu 100

Db        108 ------------------------------------------------------------ 108

Qy        101 LeuLeuCysSerValProPheValAlaGlyPheAlaValIleThrAlaAlaGlnAspVal 120

Db        108 ------------------------------------------------------------ 108

Qy        121 TrpMetLeuLeuGlyGlyArgLeuLeuThrGlyLeuAlaCysGlyValAlaSerLeuVal 140

Db        108 ------------------------------------------------------------ 108

Qy        141 AlaProValTyrIleSerGluIleAlaTyrProAlaValArgGlyLeuLeuGlySerCys 160
                    ||||||||||||||||||   |||||||||||||||||||||||||||||||||
Db        109 ------GTCTACATCTCCGAAATCRCCTACCCAGCAGTCCGGGGGTTGCTCGGCTCCTGT 162

Qy        161 ValGlnLeuMetValValValGlyIleLeuLeuAlaTyrLeuAlaGlyTrpValLeuGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        163 GTGCAGCTAATGGTCGTCGTCGGCATCCTCCTGGCCTACCTGGCAGGCTGGGTGCTGGAG 222

Qy        181 TrpArgTrpLeuAlaValLeuGlyCysValProProSerLeuMetLeuLeuLeuMetCys 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        223 TGGCGCTGGCTGGCTGTGCTGGGCTGCGTGCCCCCCTCCCTCATGCTGCTTCTCATGTGC 282

Qy        201 PheMetProGluThrProArgPheLeuLeuThrGlnHisArgArgGlnGluAlaMetAla 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        283 TTCATGCCCGAGACCCCGCGCTTCCTGCTGACTCAGCACAGGCGCCAGGAGGCCATGGCC 342

Qy        221 AlaLeuArgPheLeuTrpGlySerGluGlnGlyTrpGluAspProProIleGlyAlaGlu 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||   |||
Db        343 GCCCTGCGNTTCCTGTGGGGCTCCGAGCAGGGCTGGGAAGACCCCCCCATCGGGNCTGAG 402

Qy        241 GlnSerPheHisLeuAlaLeuLeuArgGlnProGlyIleTyrLysProPheIleIleGly 260
              |||||||||||||||||||||||||||   ||||||||||||||||||||||||||||||
Db        403 CAGAGCTTTCACCTGGCCCTGCTGCGGCANCCCGGCATCTACAAGCCCTTCATCATCGGC 462
```

```
Qy        261 ValSerLeuMetAlaPheGlnGlnLeuSerGlyValAsnAlaValMetPheTyrAlaGlu 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        463 GTCTCCCTGATGGCCTTCCAGCAGCTGTCGGGGGTCAACGCCGTCATGTTCTATGCAGAG 522

Qy        281 ThrIlePheGluGluAlaLysPheLysAspSerSerLeuAlaSerValValValGlyVal 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        523 ACCATCTTTGAAGAGGCCAAGTTCAAGGACAGCAGCCTGGCCTCGGTCGTCGTGGGTGTC 582

Qy        301 IleGlnValLeuPheThrAlaValAlaAlaLeuIleMetAspArgAlaGlyArgArgLeu 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        583 ATCCAGGTGCTGTTCACAGCTGTGGCGGCTCTCATCATGGACAGAGCAGGGCGGAGGCTG 642

Qy        321 LeuLeuValLeuSerGlyValValMetValPheSerThrSerAlaPheGlyAlaTyrPhe 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        643 CTCCTGGTCTTGTCAGGTGTGGTCATGGTGTTCAGCACGAGTGCCTTCGGCGCCTACTTC 702

Qy        341 LysLeuThrGlnGlyGlyProGlyAsnSerSerHisValAlaIleSerAlaProValSer 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        703 AAGCTGACCCAGGGTGGCCCTGGCAACTCCTCGCACGTGGCCATCTCGGCGCCTGTCTCT 762

Qy        361 AlaGlnProValAspAlaSerValGlyLeuAlaTrpLeuAlaValGlySerMetCysLeu 380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        763 GCACAGCCTGTTGATGCCAGCGTGGGGCTNGCCTGGCTGGCCGTGGGCAGCATGTGCCTC 822

Qy        381 PheIleAlaGlyPheAlaValGlyTrpGlyProIleProTrpLeuLeuMetSerGluIle 400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        823 TTCATCGCCGGCTTTGCGGTGGGCTGGGGGCCCATCCCCTGGCTCCTCATGTCAGAGATC 882

Qy        401 PheProLeuHisValLysGlyValAlaThrGlyIleCysValLeuThrAsnTrpLeuMet 420
              |||||||||||||||||||||||||||||||||   |||||||||   ||||||||||||
Db        883 TTCCCTYTGCATGTCAAGGGCGTGGCGACAGGCAWYTGCGTCCTCAMCAACTGGCTCATG 942

Qy        421 AlaPheLeuValThrLysGluPheSerSerLeuMetGluValLeuArgProTyrGlyAla 440
                 ||||||   ||||||||||||||||||||||||||||||||||||||||||||||||
Db        943 GSCTTTCTYGKGACCAAGGAGTTCAGCAGCCTCATGGAGGTCCTCAGGCCCTATGGAGCC 1002

Qy        441 PheTrpLeuAlaSerAlaPheCysIlePheSerValLeuPheThrLeuPheCysValPro 460
              ||||||||||||||||||||||||||||||||||||||||||||||||   |||   |||
Db       1003 TTCTGGCTTGCCTCCGCTTTCTGCATCTTCAGTGTCCTTTTCACTTTGGTCTGTGC-CCT 1061

Qy        461 GluThrLysGlyLysThrLeuGluGlnIleThrAlaHisPheGluGlyArg 477
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1062 GAAACTAAAGGAAAGACTCTGGAACAAATCACAGCCCATTTTGAGGGGCGA 1112

RESULT 5
US-10-169-395-29
; Sequence 29, Application US/10169395
; Publication No. US20040034192A1
; GENERAL INFORMATION:
;  APPLICANT: KATO, Seishi
;  APPLICANT:  KIMURA, Tomoko
;  TITLE OF INVENTION: HUMAN PROTEINS HAVING HYDROPHOBIC DOMAINS AND DNAs ENCODING
;  TITLE OF INVENTION:  THESE PROTEINS
;  FILE REFERENCE: 01997.015100.US
;  CURRENT APPLICATION NUMBER: US/10/169,395
;  CURRENT FILING DATE:  2002-11-29
;  PRIOR APPLICATION NUMBER: JP 2000-585
;  PRIOR FILING DATE: 2000-01-06
;  PRIOR APPLICATION NUMBER: JP 2000-588
;  PRIOR FILING DATE: 2000-01-06
;  PRIOR APPLICATION NUMBER: JP 2000-2299
;  PRIOR FILING DATE: 2000-01-11
;  PRIOR APPLICATION NUMBER: JP 2000-26862
;  PRIOR FILING DATE: 2000-02-03
;  PRIOR APPLICATION NUMBER: JP 2000-58367
;  PRIOR FILING DATE: 2000-03-03
;  PRIOR APPLICATION NUMBER: PCT/JP00/09359
;  PRIOR FILING DATE: 2000-12-28
;  NUMBER OF SEQ ID NOS: 150
; SEQ ID NO 29
;   LENGTH: 1461
;   TYPE: DNA
;   ORGANISM: Homo sapiens
```

```
;    FEATURE:
;    NAME/KEY: CDS
;    LOCATION: (61)..(849)
US-10-169-395-29

Alignment Scores:
Pred. No.:              3.34e-157     Length:         1461
Score:                  1554.50       Matches:        324
Percent Similarity:     67.85%        Conservative:   1
Best Local Similarity:  67.64%        Mismatches:     1
Query Match:            63.27%        Indels:         153
DB:                     13            Gaps:           2

US-09-886-954A-1 (1-477) x US-10-169-395-29 (1-1461)

Qy         1 MetThrProGluAspProGluGluThrGlnProLeuLeuGlyProProGlyGlySerAla 20
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        61 ATGACGCCCGAGGACCCAGAGGAAACCCAGCCGCTTCTGGGGCCTCCTGGCGGCAGCGCG 120

Qy        21 ProArgGlyArgArgValPheLeuAlaAlaPheAlaAlaAlaLeuGlyProLeuSerPhe 40
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       121 CCCCGCGGCCGCCGCGTCTTCCTCGCCGCCTTCGCCGCTGCCCTGGGCCCACTCAGCTTC 180

Qy        41 GlyPheAlaLeuGlyTyrSerSerProAlaIleProSerLeuGlnArgAlaAlaProPro 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       181 GGCTTCGCGCTCGGCTACAGCTCCCCGGCCATCCCTAGCCTGCAGCGCGCCGCGCCCCCG 240

Qy        61 AlaProArgLeuAspAspAlaAlaAlaSerTrpPheGlyAlaValValThrLeuGlyAla 80
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241 GCCCCGCGCCTGGACGACGCCGCCGCCTCCTGGTTCGGGGCTGTCGTGACCCTGGGTGCC 300

Qy        81 AlaAlaGlyGlyValLeuGlyGlyTrpLeuValAspArgAlaGlyArgLysLeuSerLeu 100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       301 GCGGCGGGGGGAGTGCTGGGCGGCTGGCTGGTGGACCGCGCCGGGCGCAAGCTGAGCCTC 360

Qy       101 LeuLeuCysSerValProPheValAlaGlyPheAlaValIleThrAlaAlaGlnAspVal 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       361 TTGCTGTGCTCCGTGCCCTTCGTGGCCGGCTTTGCCGTCATCACCGCGGCCCAGGACGTG 420

Qy       121 TrpMetLeuLeuGlyGlyArgLeuLeuThrGlyLeuAlaCysGlyValAlaSerLeuVal 140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       421 TGGATGCTGCTGGGGGGCCGCCTCCTCACCGGCCTGGCCTGCGGTGTTGCCTCCCTAGTG 480

Qy       141 AlaProValTyrIleSerGluIleAlaTyrProAlaValArgGlyLeuLeuGlySerCys 160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       481 GCCCCGGTCTACATCTCCGAAATCGCCTACCCAGCAGTCCGGGGGTTGCTCGGCTCCTGT 540

Qy       161 ValGlnLeuMetValValValGlyIleLeuLeuAlaTyrLeuAlaGlyTrpValLeuGlu 180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       541 GTGCAGCTAATGGTCGTCGTCGGCATCCTCCTGGCCTACCTGGCAGGCTGGGTGCTGGAG 600

Qy       181 TrpArgTrpLeuAlaValLeuGlyCysValProProSerLeuMetLeuLeuLeuMetCys 200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       601 TGGCGCTGGCTGGCTGTGCTGGGCTGCGTGCCCCCCTCCCTCATGCTGCTTCTCATGTGC 660

Qy       201 PheMetProGluThrProArgPheLeuLeuThrGlnHisArgArgGlnGluAla-MetAl 220
             |||||||||||||||||||||||||||||||||||||||||||||||||||||| :::
Db       661 TTCATGCCCGAGACCCCGCGCTTCCTGCTGACTCAGCACAGGCGCCAGGAGGCTGCTC-- 718

Qy       220 aAlaLeuArgPheLeuTrpGlySerGluGlnGlyTrpGluAspProProIleGlyAlaGl 240

Db       718 ------------------------------------------------------------ 718

Qy       240 uGlnSerPheHisLeuAlaLeuLeuArgGlnProGlyIleTyrLysProPheIleIleGl 260

Db       718 ------------------------------------------------------------ 718

Qy       260 yValSerLeuMetAlaPheGlnGlnLeuSerGlyValAsnAlaValMetPheTyrAlaGl 280

Db       718 ------------------------------------------------------------ 718

Qy       280 uThrIlePheGluGluAlaLysPheLysAspSerSerLeuAlaSerValValValGlyVa 300

Db       718 ------------------------------------------------------------ 718

Qy       300 lIleGlnValLeuPheThrAlaValAlaAlaLeuIleMetAspArgAlaGlyArgArgLe 320

Db       718 ------------------------------------------------------------ 718

Qy       320 uLeuLeuValLeuSerGlyValValMetValPheSerThrSerAlaPheGlyAlaTyrPh 340
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       719 ----CTGGTCTTGTCAGGTGTGGTCATGGTGTTCAGCACGAGTGCCTTCGGCGCCTACTT 774

Qy       340 eLysLeuThrGlnGlyGlyProGlyAsnSerSerHisValAlaIleSerAlaProValSe 360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       775 CAAGCTGACCCAGGGTGGCCCTGGCAACTCCTCGCACGTGGCCATCTCGGCGCCTGTCTC 834

Qy       360 rAlaGlnProValAspAlaSerValGlyLeuAlaTrpLeuAlaValGlySerMetCysLe 380
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       835 TGCACAGCCTGTTGATGCCAGCGTGGGGCTGGCCTGGCTGGCCGTGGGCAGCATGTGCCT 894

Qy       380 uPheIleAlaGly-PheAlaValGlyTrpGlyProIleProTrpLeuLeuMetSerGluI 400
             |||||||||||||
Db       895 CTTCATCGCCGGAGGTC------------------------------------------- 911

Qy       400 lePheProLeuHisValLysGlyValAlaThrGlyIleCysValLeuThrAsnTrpLeuM 420

Db       911 ------------------------------------------------------------ 911

Qy       420 etAlaPheLeuValThrLysGluPheSerSerLeuMetGluValLeuArgProTyrGlyA 440
                                                         ||||||||||||||||
Db       912 --------------------------------------------CTCAGGCCCTATGGAG 927

Qy       440 laPheTrpLeuAlaSerAlaPheCysIlePheSerValLeuPheThrLeuPheCysValP 460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       928 CCTTCTGGCTTGCCTCCGCTTTCTGCATCTTCAGTGTCCTTTTCACTTTGTTCTGTGTCC 987

Qy       460 roGluThrLysGlyLysThrLeuGluGlnIleThrAlaHisPheGluGlyArg 477
             |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       988 CTGAAACTAAAGGAAAGACTCTGGAACAAATCACAGCCCATTTTGAGGGGCGA 1040

RESULT 6
US-10-169-395-19
; Sequence 19, Application US/10169395
; Publication No. US20040034192A1
; GENERAL INFORMATION:
;  APPLICANT: KATO, Seishi
;  APPLICANT:  KIMURA, Tomoko
;  TITLE OF INVENTION: HUMAN PROTEINS HAVING HYDROPHOBIC DOMAINS AND DNAs ENCODING
;  TITLE OF INVENTION:  THESE PROTEINS
;  FILE REFERENCE: 01997.015100.US
;  CURRENT APPLICATION NUMBER: US/10/169,395
;  CURRENT FILING DATE:  2002-11-29
;  PRIOR APPLICATION NUMBER: JP 2000-585
;  PRIOR FILING DATE: 2000-01-06
;  PRIOR APPLICATION NUMBER: JP 2000-588
;  PRIOR FILING DATE: 2000-01-06
;  PRIOR APPLICATION NUMBER: JP 2000-2299
;  PRIOR FILING DATE: 2000-01-11
;  PRIOR APPLICATION NUMBER: JP 2000-26862
;  PRIOR FILING DATE: 2000-02-03
;  PRIOR APPLICATION NUMBER: JP 2000-58367
;  PRIOR FILING DATE: 2000-03-03
;  PRIOR APPLICATION NUMBER: PCT/JP00/09359
;  PRIOR FILING DATE: 2000-12-28
;  NUMBER OF SEQ ID NOS: 150
; SEQ ID NO 19
;   LENGTH: 789
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-169-395-19

Alignment Scores:
Pred. No.:              9.99e-127     Length:         789
Score:                  1268.00       Matches:        261
Percent Similarity:     71.78%        Conservative:   1
```

```
Best Local Similarity:  71.51%        Mismatches:     0
Query Match:            51.61%        Indels:       103
DB:                     13            Gaps:           1

US-09-886-954A-1 (1-477) x US-10-169-395-19 (1-789)

Qy         1 MetThrProGluAspProGluGluThrGlnProLeuLeuGlyProProGlyGlySerAla 20
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1 ATGACGCCCGAGGACCCAGAGGAAACCCAGCCGCTTCTGGGGCCTCCTGGCGGCAGCGCG 60

Qy        21 ProArgGlyArgArgValPheLeuAlaAlaPheAlaAlaAlaLeuGlyProLeuSerPhe 40
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        61 CCCCGCGGCCGCCGCGTCTTCCTCGCCGCCTTCGCCGCTGCCCTGGGCCCACTCAGCTTC 120

Qy        41 GlyPheAlaLeuGlyTyrSerSerProAlaIleProSerLeuGlnArgAlaAlaProPro 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       121 GGCTTCGCGCTCGGCTACAGCTCCCCGGCCATCCCTAGCCTGCAGCGCGCCGCGCCCCCG 180

Qy        61 AlaProArgLeuAspAspAlaAlaAlaSerTrpPheGlyAlaValValThrLeuGlyAla 80
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       181 GCCCCGCGCCTGGACGACGCCGCCGCCTCCTGGTTCGGGGCTGTCGTGACCCTGGGTGCC 240

Qy        81 AlaAlaGlyGlyValLeuGlyGlyTrpLeuValAspArgAlaGlyArgLysLeuSerLeu 100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241 GCGGCGGGGGGAGTGCTGGGCGGCTGGCTGGTGGACCGCGCCGGGCGCAAGCTGAGCCTC 300

Qy       101 LeuLeuCysSerValProPheValAlaGlyPheAlaValIleThrAlaAlaGlnAspVal 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       301 TTGCTGTGCTCCGTGCCCTTCGTGGCCGGCTTTGCCGTCATCACCGCGGCCCAGGACGTG 360

Qy       121 TrpMetLeuLeuGlyGlyArgLeuLeuThrGlyLeuAlaCysGlyValAlaSerLeuVal 140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       361 TGGATGCTGCTGGGGGGCCGCCTCCTCACCGGCCTGGCCTGCGGTGTTGCCTCCCTAGTG 420

Qy       141 AlaProValTyrIleSerGluIleAlaTyrProAlaValArgGlyLeuLeuGlySerCys 160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       421 GCCCCGGTCTACATCTCCGAAATCGCCTACCCAGCAGTCCGGGGGTTGCTCGGCTCCTGT 480

Qy       161 ValGlnLeuMetValValValGlyIleLeuLeuAlaTyrLeuAlaGlyTrpValLeuGlu 180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       481 GTGCAGCTAATGGTCGTCGTCGGCATCCTCCTGGCCTACCTGGCAGGCTGGGTGCTGGAG 540

Qy       181 TrpArgTrpLeuAlaValLeuGlyCysValProProSerLeuMetLeuLeuLeuMetCys 200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       541 TGGCGCTGGCTGGCTGTGCTGGGCTGCGTGCCCCCCTCCCTCATGCTGCTTCTCATGTGC 600

Qy       201 PheMetProGluThrProArgPheLeuLeuThrGlnHisArgArgGlnGluAla-MetAl 220
             |||||||||||||||||||||||||||||||||||||||||||||||||||||| :::
Db       601 TTCATGCCCGAGACCCCGCGCTTCCTGCTGACTCAGCACAGGCGCCAGGAGGCTGCTC-- 658

Qy       220 aAlaLeuArgPheLeuTrpGlySerGluGlnGlyTrpGluAspProProIleGlyAlaGl 240

Db       658 ------------------------------------------------------------ 658

Qy       240 uGlnSerPheHisLeuAlaLeuLeuArgGlnProGlyIleTyrLysProPheIleIleGl 260

Db       658 ------------------------------------------------------------ 658

Qy       260 yValSerLeuMetAlaPheGlnGlnLeuSerGlyValAsnAlaValMetPheTyrAlaGl 280

Db       658 ------------------------------------------------------------ 658

Qy       280 uThrIlePheGluGluAlaLysPheLysAspSerSerLeuAlaSerValValValGlyVa 300

Db       658 ------------------------------------------------------------ 658

Qy       300 lIleGlnValLeuPheThrAlaValAlaAlaLeuIleMetAspArgAlaGlyArgArgLe 320

Db       658 ------------------------------------------------------------ 658

Qy       320 uLeuLeuValLeuSerGlyValValMetValPheSerThrSerAlaPheGlyAlaTyrPh 340
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       659 ----CTGGTCTTGTCAGGTGTGGTCATGGTGTTCAGCACGAGTGCCTTCGGCGCCTACTT 714
```

```
Qy       340 eLysLeuThrGlnGlyGlyProGlyAsnSerSerHisValAlaIleSerAlaProValSe 360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       715 CAAGCTGACCCAGGGTGGCCCTGGCAACTCCTCGCACGTGGCCATCTCGGCGCCTGTCTC 774

Qy       360 rAlaGlnProVal 364
             |||||||||||||
Db       775 TGCACAGCCTGTT 787

RESULT 7
US-10-157-031-358
; Sequence 358, Application US/10157031
; Publication No. US20030108890A1
; GENERAL INFORMATION:
;  APPLICANT: Baranova, A. V.
;  APPLICANT:  Yankovsky, N. K.
;  APPLICANT:  Kozlov, A. P.
;  APPLICANT:  Lobashev, A. V.
;  APPLICANT:  Krukovskaya, L. L.
;  TITLE OF INVENTION: In silico screening for phenotype-associated expressed sequences
;  FILE REFERENCE: 2760-103
;  CURRENT APPLICATION NUMBER: US/10/157,031
;  CURRENT FILING DATE:  2002-05-30
;  NUMBER OF SEQ ID NOS: 415
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 358
;   LENGTH: 2356
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-157-031-358

Alignment Scores:
Pred. No.:               1.01e-92       Length:        2356
Score:                   958.50         Matches:       227
Percent Similarity:      59.22%         Conservative:  78
Best Local Similarity:   44.08%         Mismatches:    164
Query Match:             39.01%         Indels:        47
DB:                      15             Gaps:          12

US-09-886-954A-1 (1-477) x US-10-157-031-358 (1-2356)

Qy         3 ProGlu---AspProGluGluThrGlnProLeuLeuGly--------------------- 14
             ||||||   ||||||         :::||||||||||||
Db        12 CCCGAGAGAGACCCGGCCATGCAGGAGCCGCTGCTGGGAGCCGAGGGCCCGGACTACGAC 71

Qy        15 ---------------ProProGlyGlySerAlaProArgGlyArg--------------- 24
                            ||||||               |||   |||
Db        72 ACCTTCCCCGAGAAGCCGCCCCCGTCGCCAGGGGACAGGGCGCGGGTCGGGACCCTGCAG 131

Qy        25 ------ArgValPheLeuAlaAlaPheAlaAlaAlaLeuGlyProLeuSerPheGlyPhe 42
                   |||||||||||||||   |||||||||   ||||||      |||||||||:::
Db       132 AACAAAAGGGTGTTCCTGGCCACCTTCGCCGCAGTGCTCGGCAATTTCAGCTTTGGGTAT 191

Qy        43 AlaLeuGlyTyrSerSerProAlaIleProSerLeuGlnArgAlaAlaProProAlaPro 62
             ||||||   |||:::||||||   ||||||:::|||:::|||:::      |||
Db       192 GCCCTGGTCTACACATCCCCTGTCATCCCAGCCCTGGAGCGCTCCTTGGATCCTGACCTG 251

Qy        63 ArgLeuAspAspAlaAlaAlaSerTrpPheGlyAlaValValThrLeuGlyAlaAlaAla 82
                |||      :::   |||||||||||||||:::|||   ||||||||||||||||||
Db       252 CATCTGACCAAATCCCAGGCATCCTGGTTTGGGTCCGTGTTCACCCTGGGAGCAGCGGCC 311

Qy        83 GlyGlyValLeuGlyGlyTrpLeuValAspArgAlaGlyArgLysLeuSerLeuLeuLeu 102
             ||||||:::            |||   |||      |||||||||||||||::::::
Db       312 GGAGGCCTGAGTGCCATGATCCTCAACGACCTCCTGGGCCGGAAGCTGAGCATCATGTTC 371

Qy       103 CysSerValProPheValAlaGlyPheAlaValIleThrAlaAlaGlnAspValTrpMet 122
                :::||||||      ||||||:::|||::::::      |||      :::||||||
Db       372 TCAGCTGTGCCGTCGGCGGCCGGCTATGCGCTCATGGCGGGTGCGCACGGCCTCTGGATG 431

Qy       123 LeuLeuGlyGlyArgLeuLeuThrGlyLeuAlaCysGlyValAlaSerLeuValAlaPro 142
             ||||||   ||||||   |||||||||   |||   |||:::   :::         |||
Db       432 CTGCTGCTCGGAAGGACGCTGACGGGCTTCGCCGGGGGGCTCACAGCTGCCTGCATCCCG 491
```

```
Qy      143 ValTyrIleSerGluIleAlaTyrProAlaValArgGlyLeuLeuGlySerCysValGln 162
            ||||||:::||||||||||||   |||   |||||||||   ||||||:::      |||
Db      492 GTGTACGTGTCTGAGATTGCTCCCCCAGGCGTTCGTGGGGCTCTGGGGGCCACACCCCAG 551

Qy      163 LeuMetValValValGlyIleLeuLeuAlaTyrLeuAlaGlyTrpValLeuGluTrpArg 182
            ||||||   |||   |||   |||      |||      |||   :::|||   ||||||
Db      552 CTCATGGCAGTGTTCGGATCCCTGTCCCTCTACGCCCTTGGCCTCCTGCTGCCGTGGCGC 611

Qy      183 TrpLeuAlaValLeuGlyCysValProProSerLeuMetLeuLeuLeuMetCysPheMet 202
            ||||||||||||   |||      |||      :::|||:::||||||:::   ||||||
Db      612 TGGCTGGCTGTGGCCGGGGAGGCGCCTGTGCTCATCATGATCCTGCTGCTCAGCTTCATG 671

Qy      203 ProGluThrProArgPheLeuLeuThrGlnHisArgArgGlnGluAlaMetAlaAlaLeu 222
            |||   :::|||||||||||||||::::::   |||   :::||||||:::   ||||||
Db      672 CCCAACTCGCCGCGCTTCCTGCTCTCTCGGGGCAGGGACGAAGAGGCCCTGCGGGCGCTG 731

Qy      223 ArgPheLeuTrpGlySerGluGlnGly------TrpGluAspProProIleGly------ 238
               :::|||   |||::::::            ||||||               |||
Db      732 GCCTGGCTGCGTGGGACGGACGTCGATGTCCACTGGGAGTTCGAGCAGATCCAGGACAAC 791

Qy      239 ---AlaGluGlnSerPheHisLeu------AlaLeuLeuArgGlnProGlyIleTyrLys 255
                     ||||||      :::      |||      |||   |||   :::   :::
Db      792 GTCCGGAGACAGAGCAGCCGAGTATCGTGGGCTGAGGCACGGGCCCCCACGTGTGCCGG 851

Qy      256 ProPheIleIleGlyValSerLeuMetAlaPheGlnGlnLeuSerGlyValAsnAlaVal 275
            |||      :::   :::   :::         |||||||||:::|||:::      :::
Db      852 CCCATCACCGTGGCCTTGCTGATGCGCCTCCTGCAGCAGCTGACGGGCATCACGCCCATC 911

Qy      276 MetPheTyrAlaGluThrIlePheGluGluAlaLysPhe------------LysAspSer 291
            :::   |||   ::::::||||||:::                        ||||||
Db      912 CTGGTCTACCTGCAGTCCATCTTCGACAGCACCGCTGTCCTGCTGCCCCCCAAGGACGAC 971

Qy      292 SerLeuAlaSerValValValGlyValIleGlnValLeuPheThrAlaValAlaAlaLeu 311
                     :::   :::||||||   :::::::::|||         :::|||||||||
Db      972 ---------GCAGCCATCGTTGGGGCCGTGCGGCTCCTGTCCGTGCTGATCGCCGCCCTC 1022

Qy      312 IleMetAspArgAlaGlyArgArgLeuLeuLeuValLeuSerGlyValValMetValPhe 331
               ||||||   |||||||||::::::||||||   :::|||      :::|||
Db     1023 ACCATGGACCTCGCAGGCCGCAAGGTGCTGCTCTTCGTCTCAGCGGCCATCATGTTTGCT 1082

Qy      332 SerThrSerAlaPheGlyAlaTyrPheLysLeuThrGlnGlyGlyProGlyAsnSerSer 351
            :::            |||   |||                  ||||||         |||
Db     1083 GCCAACCTGACTCTGGGGCTGTACATCCACTTT---------GGCCCCAGGCCTCTGAGC 1133

Qy      352 HisValAlaIleSerAlaProValSer------------------AlaGlnProValAsp 365
                  :::   :::         |||                  |||||||||:::
Db     1134 CCCAACAGCACTGCGGGCCTGGAAAGCGAGTCCTGGGGGGACTTGGCGCAGCCCCTGGCA 1193

Qy      366 AlaSerValGly---LeuAlaTrpLeuAlaValGlySerMetCysLeuPheIleAlaGly 384
            |||      |||   |||         :::   :::   :::         |||||||||   |||
Db     1194 GCACCCGCTGGCTACCTCACCCTGGTGCCCCTGCTGGCCACCATGCTCTTCATCATGGGC 1253

Qy      385 PheAlaValGlyTrpGlyProIleProTrpLeuLeuMetSerGluIlePheProLeuHis 404
            :::|||||||||||||||||||||   ||||||||||||||||||:::   ||||||
Db     1254 TACGCCGTGGGCTGGGGTCCCATCACCTGGCTGCTCATGTCTGAGGTCCTGCCCCTGCGT 1313

Qy      405 ValLysGlyValAlaThrGlyIleCysValLeuThrAsnTrpLeuMetAlaPheLeuVal 424
               :::|||||||||:::|||:::|||||||||   :::||||||   ||||||::::::
Db     1314 GCCCGTGGCGTGGCCTCAGGGCTCTGCGTGCTGGCCAGCTGGCTCACCGCCTTCGTCCTC 1373

Qy      425 ThrLysGluPheSerSerLeuMetGluValLeuArgProTyrGlyAlaPheTrpLeuAla 444
            ||||||   |||      :::   |||   |||||||||   |||||||||   |||
Db     1374 ACCAAGTCCTTCCTGCCAGTGGT-GAGCACCTTCGGCCTCCAGGTGCCTTTCTTCTTCTT 1432

Qy      445 -SerAlaPheCysIlePheSerValLeuPheThrLeuPheCysValProGluThrLysGl 464
             :::|||   |||:::   |||::::::||||||      ||||||||||||||||||||
Db     1433 CGCGGCCATCTGCTTGGTGAGCCTGGTGTTCACAGGCTGCTGTGTGCCCGAGACCAAGGG 1492

Qy      464 yLysThrLeuGluGlnIleThrAlaHisPheGlu---GlyArg 477
            |::::::||||||||||||   :::   |||      ||||||
Db     1493 ACGGTCCCTGGAGCAGATCGAGTCCTTCTTCCGCACGGGGAGA 1535

RESULT 8
```

```
US-10-342-887-1685
; Sequence 1685, Application US/10342887
; Publication No. US20040058340A1
; GENERAL INFORMATION:
;  APPLICANT: Dai, Hongyue
;  APPLICANT:  He, Yudong
;  APPLICANT:  Linsley, Peter S.
;  APPLICANT:  Mao, Mao
;  APPLICANT:  Roberts, Christopher J.
;  APPLICANT:  Van 't Veer, Laura Johanna
;  APPLICANT:  Van de Vijver, Marc J.
;  APPLICANT:  Bernards, Rene
;  TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
;  FILE REFERENCE: 9301-188-999
;  CURRENT APPLICATION NUMBER: US/10/342,887
;  CURRENT FILING DATE:  2003-01-15
;  PRIOR APPLICATION NUMBER: 60/298,918
;  PRIOR FILING DATE: 2001-06-18
;  PRIOR APPLICATION NUMBER: 60/380,710
;  PRIOR FILING DATE: 2002-05-14
;  PRIOR APPLICATION NUMBER: 10/172,118
;  PRIOR FILING DATE: 2002-06-14
;  NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 1685
;   LENGTH: 2487
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-342-887-1685

Alignment Scores:
Pred. No.:               1.09e-92       Length:         2487
Score:                   958.50         Matches:        227
Percent Similarity:      59.22%         Conservative:   78
Best Local Similarity:   44.08%         Mismatches:     164
Query Match:             39.01%         Indels:         47
DB:                      13             Gaps:           12

US-09-886-954A-1 (1-477) x US-10-342-887-1685 (1-2487)

Qy        3 ProGlu---AspProGluGluThrGlnProLeuLeuGly--------------------- 14
            ||||||   ||||||            :::||||||||||||
Db       20 CCCGAGAGAGACCCGGCCATGCAGGAGCCGCTGCTGGGAGCCGAGGGCCCGGACTACGAC 79

Qy       15 ---------------ProProGlyGlySerAlaProArgGlyArg--------------- 24
                           ||||||                  |||   |||
Db       80 ACCTTCCCCGAGAAGCCGCCCCCGTCGCCAGGGGACAGGGCGCGGGTCGGGACCCTGCAG 139

Qy       25 ------ArgValPheLeuAlaAlaPheAlaAlaAlaLeuGlyProLeuSerPheGlyPhe 42
                  |||||||||||||||   |||||||||   ||||||      |||||||||:::
Db      140 AACAAAAGGGTGTTCCTGGCCACCTTCGCCGCAGTGCTCGGCAATTTCAGCTTTGGGTAT 199

Qy       43 AlaLeuGlyTyrSerSerProAlaIleProSerLeuGlnArgAlaAlaProProAlaPro 62
            ||||||   |||:::||||||   ||||||:::|||:::|||:::         |||
Db      200 GCCCTGGTCTACACATCCCCTGTCATCCCAGCCCTGGAGCGCTCCTTGGATCCTGACCTG 259

Qy       63 ArgLeuAspAspAlaAlaAlaSerTrpPheGlyAlaValValThrLeuGlyAlaAlaAla 82
               |||      :::   |||||||||||||||:::|||   ||||||||||||||||||
Db      260 CATCTGACCAAATCCCAGGCATCCTGGTTTGGGTCCGTGTTCACCCTGGGAGCAGCGGCC 319

Qy       83 GlyGlyValLeuGlyGlyTrpLeuValAspArgAlaGlyArgLysLeuSerLeuLeuLeu 102
            ||||||:::            |||   |||         |||||||||||||||::::::
Db      320 GGAGGCCTGAGTGCCATGATCCTCAACGACCTCCTGGGCCGGAAGCTGAGCATCATGTTC 379

Qy      103 CysSerValProPheValAlaGlyPheAlaValIleThrAlaAlaGlnAspValTrpMet 122
               :::||||||      ||||||:::|||::::::      |||      :::||||||
Db      380 TCAGCTGTGCCGTCGGCGGCCGGCTATGCGCTCATGGCGGGTGCGCACGGCCTCTGGATG 439

Qy      123 LeuLeuGlyGlyArgLeuLeuThrGlyLeuAlaCysGlyValAlaSerLeuValAlaPro 142
            ||||||   ||||||   |||||||||   |||   |||:::   :::         |||
Db      440 CTGCTGCTCGGAAGGACGCTGACGGGCTTCGCCGGGGGGCTCACAGCTGCCTGCATCCCG 499

Qy      143 ValTyrIleSerGluIleAlaTyrProAlaValArgGlyLeuLeuGlySerCysValGln 162
            ||||||:::||||||||||||   |||   |||||||||   ||||||:::      |||
```

```
Db        500 GTGTACGTGTCTGAGATTGCTCCCCCAGGCGTTCGTGGGGCTCTGGGGGCCACACCCCAG 559

Qy        163 LeuMetValValValGlyIleLeuLeuAlaTyrLeuAlaGlyTrpValLeuGluTrpArg 182
              ||||||   |||   |||   |||      |||      |||   :::|||   ||||||
Db        560 CTCATGGCAGTGTTCGGATCCCTGTCCCTCTACGCCCTTGGCCTCCTGCTGCCGTGGCGC 619

Qy        183 TrpLeuAlaValLeuGlyCysValProProSerLeuMetLeuLeuLeuMetCysPheMet 202
              ||||||||||||   |||      |||      :::|||:::||||||:::   ||||||
Db        620 TGGCTGGCTGTGGCCGGGGAGGCGCCTGTGCTCATCATGATCCTGCTGCTCAGCTTCATG 679

Qy        203 ProGluThrProArgPheLeuLeuThrGlnHisArgArgGlnGluAlaMetAlaAlaLeu 222
              |||   :::|||||||||||||||::::::   |||   :::||||||:::   ||||||
Db        680 CCCAACTCGCCGCGCTTCCTGCTCTCGGGGCAGGGACGAAGAGGCCCTGCGGGCGCTG 739

Qy        223 ArgPheLeuTrpGlySerGluGlnGly------TrpGluAspProProIleGly------ 238
                 :::|||   |||::::::            ||||||         |||
Db        740 GCCTGGCTGCGTGGGACGGACGTCGATGTCCACTGGGAGTTCGAGCAGATCCAGGACAAC 799

Qy        239 ---AlaGluGlnSerPheHisLeu------AlaLeuLeuArgGlnProGlyIleTyrLys 255
                       ||||||      :::      |||      |||   |||   :::   :::
Db        800 GTCCGGAGACAGAGCAGCCGAGTATCGTGGGCTGAGGCACGGGCCCCACACGTGTGCCGG 859

Qy        256 ProPheIleIleGlyValSerLeuMetAlaPheGlnGlnLeuSerGlyValAsnAlaVal 275
              |||      :::   :::   :::         |||||||||:::|||:::      :::
Db        860 CCCATCACCGTGGCCTTGCTGATGCGCCTCCTGCAGCAGCTGACGGGCATCACGCCCATC 919

Qy        276 MetPheTyrAlaGluThrIlePheGluGluAlaLysPhe------------LysAspSer 291
              :::   |||   ::::::||||||:::                        ||||||
Db        920 CTGGTCTACCTGCAGTCCATCTTCGACAGCACCGCTGTCCTGCTGCCCCCCAAGGACGAC 979

Qy        292 SerLeuAlaSerValValValGlyValIleGlnValLeuPheThrAlaValAlaAlaLeu 311
                       :::   :::||||||   :::::::::|||         :::|||||||||
Db        980 ---------GCAGCCATCGTTGGGGCCGTGCGGCTCCTGTCCGTGCTGATCGCCGCCCTC 1030

Qy        312 IleMetAspArgAlaGlyArgArgLeuLeuLeuValLeuSerGlyValValMetValPhe 331
                 ||||||   |||||||||::::::||||||   :::|||      :::|||
Db       1031 ACCATGGACCTCGCAGGCCGCAAGGTGCTGCTCTTCGTCTCAGCGGCCATCATGTTTGCT 1090

Qy        332 SerThrSerAlaPheGlyAlaTyrPheLysLeuThrGlnGlyGlyProGlyAsnSerSer 351
              :::            |||   |||                  ||||||         |||
Db       1091 GCCAACCTGACTCTGGGGCTGTACATCCACTTT---------GGCCCCAGGCCTCTGAGC 1141

Qy        352 HisValAlaIleSerAlaProValSer------------------AlaGlnProValAsp 365
                    :::   :::         |||                  |||||||||:::
Db       1142 CCCAACAGCACTGCGGGCCTGGAAAGCGAGTCCTGGGGGGACTTGGCGCAGCCCCTGGCA 1201

Qy        366 AlaSerValGly---LeuAlaTrpLeuAlaValGlySerMetCysLeuPheIleAlaGly 384
              |||      |||   |||      :::   :::   :::      |||||||||   |||
Db       1202 GCACCCGCTGGCTACCTCACCCTGGTGCCCCTGCTGGCCACCATGCTCTTCATCATGGGC 1261

Qy        385 PheAlaValGlyTrpGlyProIleProTrpLeuLeuMetSerGluIlePheProLeuHis 404
              :::|||||||||||||||||||||   ||||||||||||||||||:::   ||||||
Db       1262 TACGCCGTGGGCTGGGGTCCCATCACCTGGCTGCTCATGTCTGAGGTCCTGCCCCTGCGT 1321

Qy        405 ValLysGlyValAlaThrGlyIleCysValLeuThrAsnTrpLeuMetAlaPheLeuVal 424
                 :::|||||||||:::|||:::|||||||||   :::||||||   ||||||::::::
Db       1322 GCCCGTGGCGTGGCCTCAGGGCTCTGCGTGCTGGCCAGCTGGCTCACCGCCTTCGTCCTC 1381

Qy        425 ThrLysGluPheSerSerLeuMetGluValLeuArgProTyrGlyAlaPheTrpLeuAla 444
              ||||||   |||      :::   |||   |||||||||   |||||||||   |||
Db       1382 ACCAAGTCCTTCCTGCCAGTGGT-GAGCACCTTCGGCCTCCAGGTGCCTTTCTTCTTCTT 1440

Qy        445 -SerAlaPheCysIlePheSerValLeuPheThrLeuPheCysValProGluThrLysGl 464
               :::|||   |||:::   |||::::::||||||      ||||||||||||||||||||
Db       1441 CGCGGCCATCTGCTTGGTGAGCCTGGTGTTCACAGGCTGCTGTGTGCCCGAGACCAAGGG 1500

Qy        464 yLysThrLeuGluGlnIleThrAlaHisPheGlu---GlyArg 477
              |::::::||||||||||||   :::   |||      ||||||
Db       1501 ACGGTCCCTGGAGCAGATCGAGTCCTTCTTCCGCATGGGGAGA 1543

RESULT 9
US-10-172-118-1685
; Sequence 1685, Application US/10172118
```

```
; Publication No. US20030224374A1
; GENERAL INFORMATION:
;  APPLICANT: Dai, Hongyue
;  APPLICANT:  He, Yudong
;  APPLICANT:  Linsley, Peter
;  APPLICANT:  Mao, Mao
;  APPLICANT:  Roberts, Chris
;  APPLICANT:  Van 't Veer, Laura
;  APPLICANT:  Van de Vijver, Marc
;  APPLICANT:  Bernards, Rene
;  TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
;  FILE REFERENCE: 9301-175-999
;  CURRENT APPLICATION NUMBER: US/10/172,118
;  CURRENT FILING DATE:  2002-06-14
;  PRIOR APPLICATION NUMBER: 60/380,770
;  PRIOR FILING DATE: 2002-05-14
;  NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 1685
;   LENGTH: 2487
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;  PUBLICATION INFORMATION:
;   DATABASE ACCESSION NUMBER: NM_017585
;   DATABASE ENTRY DATE: 2001-06-18
US-10-172-118-1685

Alignment Scores:
Pred. No.:                1.09e-92      Length:         2487
Score:                    958.50        Matches:        227
Percent Similarity:       59.22%        Conservative:   78
Best Local Similarity:    44.08%        Mismatches:     164
Query Match:              39.01%        Indels:         47
DB:                       13            Gaps:           12

US-09-886-954A-1 (1-477) x US-10-172-118-1685 (1-2487)

Qy          3 ProGlu---AspProGluGluThrGlnProLeuLeuGly--------------------- 14
              ||||||   ||||||         :::||||||||||||
Db         20 CCCGAGAGAGACCCGGCCATGCAGGAGCCGCTGCTGGGAGCCGAGGGCCCGGACTACGAC 79

Qy         15 ---------------ProProGlyGlySerAlaProArgGlyArg--------------- 24
                             ||||||               |||   |||
Db         80 ACCTTCCCCGAGAAGCCGCCCCCGTCGCCAGGGGACAGGGCGCGGGTCGGGACCCTGCAG 139

Qy         25 ------ArgValPheLeuAlaAlaPheAlaAlaAlaLeuGlyProLeuSerPheGlyPhe 42
                    |||||||||||||||   |||||||||   ||||||      |||||||||:::
Db        140 AACAAAAGGGTGTTCCTGGCCACCTTCGCCGCAGTGCTCGGCAATTTCAGCTTTGGGTAT 199

Qy         43 AlaLeuGlyTyrSerSerProAlaIleProSerLeuGlnArgAlaAlaProProAlaPro 62
              ||||||   |||:::||||||   ||||||:::|||:::|||:::      |||
Db        200 GCCCTGGTCTACACATCCCCTGTCATCCCAGCCCTGGAGCGCTCCTTGGATCCTGACCTG 259

Qy         63 ArgLeuAspAspAlaAlaAlaSerTrpPheGlyAlaValValThrLeuGlyAlaAlaAla 82
                 |||      :::   |||||||||||||||:::|||   ||||||||||||||||||
Db        260 CATCTGACCAAATCCCAGGCATCCTGGTTTGGGTCCGTGTTCACCCTGGGAGCAGCGGCC 319

Qy         83 GlyGlyValLeuGlyGlyTrpLeuValAspArgAlaGlyArgLysLeuSerLeuLeuLeu 102
              ||||||:::            |||   |||      |||||||||||||||::::::
Db        320 GGAGGCCTGAGTGCCATGATCCTCAACGACCTCCTGGGCCGGAAGCTGAGCATCATGTTC 379

Qy        103 CysSerValProPheValAlaGlyPheAlaValIleThrAlaAlaGlnAspValTrpMet 122
                 :::||||||      ||||||:::|||::::::      |||      :::||||||
Db        380 TCAGCTGTGCCGTCGGCGGCCGGCTATGCGCTCATGGCGGGTGCGCACGGCCTCTGGATG 439

Qy        123 LeuLeuGlyGlyArgLeuLeuThrGlyLeuAlaCysGlyValAlaSerLeuValAlaPro 142
              ||||||   ||||||   |||||||||   |||   |||:::   :::         |||
Db        440 CTGCTGCTCGGAAGGACGCTGACGGGCTTCGCCGGGGGGCTCACAGCTGCCTGCATCCCG 499

Qy        143 ValTyrIleSerGluIleAlaTyrProAlaValArgGlyLeuLeuGlySerCysValGln 162
              ||||||:::||||||||||||   |||   |||||||||   ||||||:::      |||
Db        500 GTGTACGTGTCTGAGATTGCTCCCCCAGGCGTTCGTGGGGCTCTGGGGGCCACACCCCAG 559

Qy        163 LeuMetValValValGlyIleLeuLeuAlaTyrLeuAlaGlyTrpValLeuGluTrpArg 182
```

```
                ||||||   |||   |||   |||      |||      |||   :::|||   ||||||
Db          560 CTCATGGCAGTGTTCGGATCCCTGTCCCTCTACGCCCTTGGCCTCCTGCTGCCGTGGCGC 619

Qy          183 TrpLeuAlaValLeuGlyCysValProProSerLeuMetLeuLeuLeuMetCysPheMet 202
                ||||||||||||   |||      |||      :::|||:::||||||:::   ||||||
Db          620 TGGCTGGCTGTGGCCGGGGAGGCGCCTGTGCTCATCATGATCCTGCTGCTCAGCTTCATG 679

Qy          203 ProGluThrProArgPheLeuLeuThrGlnHisArgArgGlnGluAlaMetAlaAlaLeu 222
                |||   :::|||||||||||||||::::::   |||   :::||||||:::   ||||||
Db          680 CCCAACTCGCCGCGCTTCCTGCTCTCGGGGCAGGGACGAAGAGGCCCTGCGGGCGCTG 739

Qy          223 ArgPheLeuTrpGlySerGluGlnGly------TrpGluAspProProIleGly------ 238
                   :::|||   |||::::::            ||||||            |||
Db          740 GCCTGGCTGCGTGGGACGGACGTCGATGTCCACTGGGAGTTCGAGCAGATCCAGGACAAC 799

Qy          239 ---AlaGluGlnSerPheHisLeu------AlaLeuLeuArgGlnProGlyIleTyrLys 255
                         ||||||      :::      |||      |||   |||   :::   :::
Db          800 GTCCGGAGACAGAGCAGCCGAGTATCGTGGGCTGAGGCACGGGCCCCACACGTGTGCCGG 859

Qy          256 ProPheIleIleGlyValSerLeuMetAlaPheGlnGlnLeuSerGlyValAsnAlaVal 275
                |||      :::   :::   :::         |||||||||:::|||:::      :::
Db          860 CCCATCACCGTGGCCTTGCTGATGCGCCTCCTGCAGCAGCTGACGGGCATCACGCCCATC 919

Qy          276 MetPheTyrAlaGluThrIlePheGluGluAlaLysPhe------------LysAspSer 291
                :::   |||   ::::::||||||:::                  ||||||
Db          920 CTGGTCTACCTGCAGTCCATCTTCGACAGCACCGCTGTCCTGCTGCCCCCCAAGGACGAC 979

Qy          292 SerLeuAlaSerValValValGlyValIleGlnValLeuPheThrAlaValAlaAlaLeu 311
                         :::   :::||||||   :::::::::|||         :::|||||||||
Db          980 ---------GCAGCCATCGTTGGGGCCGTGCGGCTCCTGTCCGTGCTGATCGCCGCCCTC 1030

Qy          312 IleMetAspArgAlaGlyArgArgLeuLeuLeuValLeuSerGlyValValMetValPhe 331
                   ||||||   |||||||||::::::||||||   :::|||      :::|||
Db         1031 ACCATGGACCTCGCAGGCCGCAAGGTGCTGCTCTTCGTCTCAGCGGCCATCATGTTTGCT 1090

Qy          332 SerThrSerAlaPheGlyAlaTyrPheLysLeuThrGlnGlyGlyProGlyAsnSerSer 351
                :::            |||   |||                  ||||||         |||
Db         1091 GCCAACCTGACTCTGGGGCTGTACATCCACTTT---------GGCCCCAGGCCTCTGAGC 1141

Qy          352 HisValAlaIleSerAlaProValSer------------------AlaGlnProValAsp 365
                      :::   :::         |||                  |||||||||:::
Db         1142 CCCAACAGCACTGCGGGCCTGGAAAGCGAGTCCTGGGGGGACTTGGCGCAGCCCCTGGCA 1201

Qy          366 AlaSerValGly---LeuAlaTrpLeuAlaValGlySerMetCysLeuPheIleAlaGly 384
                |||      |||   |||      :::   :::   :::      |||||||||   |||
Db         1202 GCACCCGCTGGCTACCTCACCCTGGTGCCCCTGCTGGCCACCATGCTCTTCATCATGGGC 1261

Qy          385 PheAlaValGlyTrpGlyProIleProTrpLeuLeuMetSerGluIlePheProLeuHis 404
                :::|||||||||||||||||||||   ||||||||||||||||||:::   ||||||
Db         1262 TACGCCGTGGGCTGGGGTCCCATCACCTGGCTGCTCATGTCTGAGGTCCTGCCCCTGCGT 1321

Qy          405 ValLysGlyValAlaThrGlyIleCysValLeuThrAsnTrpLeuMetAlaPheLeuVal 424
                   :::|||||||||:::|||:::|||||||||   :::||||||   ||||||::::::
Db         1322 GCCCGTGGCGTGGCCTCAGGGCTCTGCGTGCTGGCCAGCTGGCTCACCGCCTTCGTCCTC 1381

Qy          425 ThrLysGluPheSerSerLeuMetGluValLeuArgProTyrGlyAlaPheTrpLeuAla 444
                ||||||   |||         :::   |||   |||||||||   |||||||||   |||
Db         1382 ACCAAGTCCTTCCTGCCAGTGGT-GAGCACCTTCGGCCTCCAGGTGCCTTTCTTCTTCTT 1440

Qy          445 -SerAlaPheCysIlePheSerValLeuPheThrLeuPheCysValProGluThrLysGl 464
                 :::|||   |||:::   |||::::::||||||      ||||||||||||||||||||
Db         1441 CGCGGCCATCTGCTTGGTGAGCCTGGTGTTCACAGGCTGCTGTGTGCCCGAGACCAAGGG 1500

Qy          464 yLysThrLeuGluGlnIleThrAlaHisPheGlu---GlyArg 477
                |::::::||||||||||||   :::   |||      ||||||
Db         1501 ACGGTCCCTGGAGCAGATCGAGTCCTTCTTCCGCATGGGGAGA 1543

RESULT 10
US-10-755-889-109
; Sequence 109, Application US/10755889
; Publication No. US20040171823A1
; GENERAL INFORMATION:
;  APPLICANT: Bristol-Myers Squibb Company
;  TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ASSOCIATED WITH THE NF-kB
;  TITLE OF INVENTION:  PATHWAY
;  FILE REFERENCE: D0284 NP
;  CURRENT APPLICATION NUMBER: US/10/755,889
;  CURRENT FILING DATE:  2004-01-13
;  PRIOR APPLICATION NUMBER: U.S. 60/440,068
;  PRIOR FILING DATE: 2003-01-14
;  PRIOR APPLICATION NUMBER: U.S. 60/469,757
;  PRIOR FILING DATE: 2003-05-12
;  NUMBER OF SEQ ID NOS: 823
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 109
;   LENGTH: 2487
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-755-889-109

Alignment Scores:
Pred. No.:                1.09e-92       Length:        2487
Score:                    958.50         Matches:       227
Percent Similarity:       59.22%         Conservative:  78
Best Local Similarity:    44.08%         Mismatches:    164
Query Match:              39.01%         Indels:        47
DB:                       17             Gaps:          12

US-09-886-954A-1 (1-477) x US-10-755-889-109 (1-2487)

Qy            3 ProGlu---AspProGluGluThrGlnProLeuLeuGly--------------------- 14
                ||||||   ||||||            :::||||||||||||
Db           20 CCCGAGAGAGACCCGGCCATGCAGGAGCCGCTGCTGGGAGCCGAGGGCCCGGACTACGAC 79

Qy           15 ---------------ProProGlyGlySerAlaProArgGlyArg--------------- 24
                               ||||||                  |||   |||
Db           80 ACCTTCCCCGAGAAGCCGCCCCCGTCGCCAGGGGACAGGGCGCGGGTCGGGACCCTGCAG 139

Qy           25 ------ArgValPheLeuAlaAlaPheAlaAlaAlaLeuGlyProLeuSerPheGlyPhe 42
                      |||||||||||||||   |||||||||   ||||||      |||||||||:::
Db          140 AACAAAAGGGTGTTCCTGGCCACCTTCGCCGCAGTGCTCGGCAATTTCAGCTTTGGGTAT 199

Qy           43 AlaLeuGlyTyrSerSerProAlaIleProSerLeuGlnArgAlaAlaProProAlaPro 62
                ||||||   |||:::||||||   ||||||:::|||:::|||:::      |||
Db          200 GCCCTGGTCTACACATCCCCTGTCATCCCAGCCCTGGAGCGCTCCTTGGATCCTGACCTG 259

Qy           63 ArgLeuAspAspAlaAlaAlaSerTrpPheGlyAlaValValThrLeuGlyAlaAlaAla 82
                   |||      :::   |||||||||||||||:::|||   ||||||||||||||||||
Db          260 CATCTGACCAAATCCCAGGCATCCTGGTTTGGGTCCGTGTTCACCCTGGGAGCAGCGGCC 319

Qy           83 GlyGlyValLeuGlyGlyTrpLeuValAspArgAlaGlyArgLysLeuSerLeuLeuLeu 102
                ||||||:::            |||   |||         |||||||||||||||::::::
Db          320 GGAGGCCTGAGTGCCATGATCCTCAACGACCTCCTGGGCCGGAAGCTGAGCATCATGTTC 379

Qy          103 CysSerValProPheValAlaGlyPheAlaValIleThrAlaAlaGlnAspValTrpMet 122
                   :::||||||      ||||||:::|||::::::      |||      :::||||||
Db          380 TCAGCTGTGCCGTCGGCGGCCGGCTATGCGCTCATGGCGGGTGCGCACGGCCTCTGGATG 439

Qy          123 LeuLeuGlyGlyArgLeuLeuThrGlyLeuAlaCysGlyValAlaSerLeuValAlaPro 142
                ||||||   ||||||   |||||||||   |||   |||:::   :::         |||
Db          440 CTGCTGCTCGGAAGGACGCTGACGGGCTTCGCCGGGGGGCTCACAGCTGCCTGCATCCCG 499

Qy          143 ValTyrIleSerGluIleAlaTyrProAlaValArgGlyLeuLeuGlySerCysValGln 162
                ||||||:::||||||||||||   |||   |||||||||   ||||||:::      |||
Db          500 GTGTACGTGTCTGAGATTGCTCCCCCAGGCGTTCGTGGGGCTCTGGGGGCCACACCCCAG 559

Qy          163 LeuMetValValValGlyIleLeuLeuAlaTyrLeuAlaGlyTrpValLeuGluTrpArg 182
                ||||||   |||   |||   |||      |||      |||   :::|||   ||||||
Db          560 CTCATGGCAGTGTTCGGATCCCTGTCCCTCTACGCCCTTGGCCTCCTGCTGCCGTGGCGC 619

Qy          183 TrpLeuAlaValLeuGlyCysValProProSerLeuMetLeuLeuLeuMetCysPheMet 202
                ||||||||||||   |||      |||      :::|||:::||||||:::   ||||||
Db          620 TGGCTGGCTGTGGCCGGGGAGGCGCCTGTGCTCATCATGATCCTGCTGCTCAGCTTCATG 679

Qy          203 ProGluThrProArgPheLeuLeuThrGlnHisArgArgGlnGluAlaMetAlaAlaLeu 222
                |||   :::|||||||||||||||::::::   |||   :::||||||:::   ||||||
```

```
Db         680 CCCAACTCGCCGCGCTTCCTGCTCTCTCGGGGCAGGGACGAAGAGGCCCTGCGGGCGCTG 739

Qy         223 ArgPheLeuTrpGlySerGluGlnGly------TrpGluAspProProIleGly------ 238
                  :::|||   |||::::::            ||||||         |||
Db         740 GCCTGGCTGCGTGGGACGGACGTCGATGTCCACTGGGAGTTCGAGCAGATCCAGGACAAC 799

Qy         239 ---AlaGluGlnSerPheHisLeu------AlaLeuLeuArgGlnProGlyIleTyrLys 255
                        ||||||      :::      |||      |||   |||   :::   :::
Db         800 GTCCGGAGACAGAGCAGCCGAGTATCGTGGGCTGAGGCACGGGCCCCACACGTGTGCCGG 859

Qy         256 ProPheIleIleGlyValSerLeuMetAlaPheGlnGlnLeuSerGlyValAsnAlaVal 275
               |||      :::   :::   :::         |||||||||:::|||:::      :::
Db         860 CCCATCACCGTGGCCTTGCTGATGCGCCTCCTGCAGCAGCTGACGGGCATCACGCCCATC 919

Qy         276 MetPheTyrAlaGluThrIlePheGluGluAlaLysPhe------------LysAspSer 291
               :::   |||   ::::::||||||:::                        ||||||
Db         920 CTGGTCTACCTGCAGTCCATCTTCGACAGCACCGCTGTCCTGCTGCCCCCCAAGGACGAC 979

Qy         292 SerLeuAlaSerValValValGlyValIleGlnValLeuPheThrAlaValAlaAlaLeu 311
                        :::   :::||||||   :::::::::|||         :::|||||||||
Db         980 ---------GCAGCCATCGTTGGGGCCGTGCGGCTCCTGTCCGTGCTGATCGCCGCCCTC 1030

Qy         312 IleMetAspArgAlaGlyArgArgLeuLeuLeuValLeuSerGlyValValMetValPhe 331
                  ||||||   |||||||||::::::||||||   :::|||      :::|||
Db        1031 ACCATGGACCTCGCAGGCCGCAAGGTGCTGCTCTTCGTCTCAGCGGCCATCATGTTTGCT 1090

Qy         332 SerThrSerAlaPheGlyAlaTyrPheLysLeuThrGlnGlyGlyProGlyAsnSerSer 351
               :::            |||   |||                  ||||||         |||
Db        1091 GCCAACCTGACTCTGGGGCTGTACATCCACTTT---------GGCCCCAGGCCTCTGAGC 1141

Qy         352 HisValAlaIleSerAlaProValSer------------------AlaGlnProValAsp 365
                     :::   :::         |||                  |||||||||:::
Db        1142 CCCAACAGCACTGCGGGCCTGGAAAGCGAGTCCTGGGGGGACTTGGCGCAGCCCCTGGCA 1201

Qy         366 AlaSerValGly---LeuAlaTrpLeuAlaValGlySerMetCysLeuPheIleAlaGly 384
               |||      |||   |||      :::   :::   :::      |||||||||   |||
Db        1202 GCACCCGCTGGCTACCTCACCCTGGTGCCCCTGCTGGCCACCATGCTCTTCATCATGGGC 1261

Qy         385 PheAlaValGlyTrpGlyProIleProTrpLeuLeuMetSerGluIlePheProLeuHis 404
               :::|||||||||||||||||||||   ||||||||||||||||||:::   ||||||
Db        1262 TACGCCGTGGGCTGGGGTCCCATCACCTGGCTGCTCATGTCTGAGGTCCTGCCCCTGCGT 1321

Qy         405 ValLysGlyValAlaThrGlyIleCysValLeuThrAsnTrpLeuMetAlaPheLeuVal 424
                  :::|||||||||:::|||:::|||||||||   :::||||||   ||||||::::::
Db        1322 GCCCGTGGCGTGGCCTCAGGGCTCTGCGTGCTGGCCAGCTGGCTCACCGCCTTCGTCCTC 1381

Qy         425 ThrLysGluPheSerSerLeuMetGluValLeuArgProTyrGlyAlaPheTrpLeuAla 444
               ||||||   |||      :::   |||   |||||||||   |||||||||   |||
Db        1382 ACCAAGTCCTTCCTGCCAGTGGT-GAGCACCTTCGGCCTCCAGGTGCCTTTCTTCTTCTT 1440

Qy         445 -SerAlaPheCysIlePheSerValLeuPheThrLeuPheCysValProGluThrLysGl 464
                :::|||   |||:::   |||::::::||||||      ||||||||||||||||||||
Db        1441 CGCGGCCATCTGCTTGGTGAGCCTGGTGTTCACAGGCTGCTGTGTGCCCGAGACCAAGGG 1500

Qy         464 yLysThrLeuGluGlnIleThrAlaHisPheGlu---GlyArg 477
               |::::::||||||||||||   :::   |||      ||||||
Db        1501 ACGGTCCCTGGAGCAGATCGAGTCCTTCTTCCGCATGGGGAGA 1543

RESULT 11
US-09-814-353-19608
; Sequence 19608, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
;  APPLICANT: Lee, John
;  APPLICANT:  Thompson, Pamela
;  APPLICANT:  Lillie, James
;  TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
;  TITLE OF INVENTION:  IDENTIFICATION, ASSESSMENT, PREVENTION, AND
;  TITLE OF INVENTION:  THERAPY OF OVARIAN CANCER
;  FILE REFERENCE: MRI-006B
;  CURRENT APPLICATION NUMBER: US/09/814,353
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: US 60/191,031
```

```
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: US 60/207,124
;  PRIOR FILING DATE: 2000-05-25
;  PRIOR APPLICATION NUMBER: US 60/211,940
;  PRIOR FILING DATE: 2000-06-15
;  PRIOR APPLICATION NUMBER: US 60/216,820
;  PRIOR FILING DATE: 2000-07-07
;  PRIOR APPLICATION NUMBER: US 60/220,661
;  PRIOR FILING DATE: 2000-07-25
;  PRIOR APPLICATION NUMBER: US 60/257,672
;  PRIOR FILING DATE: 2000-12-21
;  NUMBER OF SEQ ID NOS: 22037
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 19608
;   LENGTH: 2696
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 1, 2525, 2526, 2527, 2528, 2529, 2530, 2531, 2532, 2533,
;   LOCATION: 2534, 2535, 2536, 2537, 2538, 2539, 2540, 2541, 2542, 2543,
;   LOCATION: 2544, 2545, 2546, 2547, 2548, 2549, 2550, 2551, 2552, 2553,
;   LOCATION: 2554, 2555, 2556, 2557, 2558, 2559, 2560, 2561, 2562
;   OTHER INFORMATION: n = A,T,C or G
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 2563, 2564, 2565, 2566, 2567, 2568, 2569, 2570, 2571, 2572,
;   LOCATION: 2573, 2574, 2575, 2576, 2577, 2578, 2579, 2580, 2581, 2582,
;   LOCATION: 2583, 2584, 2585, 2586, 2587, 2588, 2589, 2590, 2591, 2592,
;   LOCATION: 2593, 2594, 2595, 2596, 2597, 2598, 2599, 2600, 2601
;   OTHER INFORMATION: n = A,T,C or G
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 2602, 2603, 2604, 2605, 2606, 2607, 2608, 2609, 2610, 2611,
;   LOCATION: 2612, 2613, 2614, 2615, 2616, 2617, 2618, 2619, 2620, 2621,
;   LOCATION: 2622, 2623, 2624, 2625, 2626, 2627, 2628, 2683, 2684, 2685,
;   LOCATION: 2686, 2687, 2688, 2689, 2690, 2691, 2692, 2693, 2694
;   OTHER INFORMATION: n = A,T,C or G
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 2695, 2696
;   OTHER INFORMATION: n = A,T,C or G
US-09-814-353-19608

Alignment Scores:
Pred. No.:                1.23e-92        Length:         2696
Score:                    958.50          Matches:        227
Percent Similarity:       59.22%          Conservative:   78
Best Local Similarity:    44.08%          Mismatches:     164
Query Match:              39.01%          Indels:         47
DB:                       10              Gaps:           12

US-09-886-954A-1 (1-477) x US-09-814-353-19608 (1-2696)

Qy           3 ProGlu---AspProGluGluThrGlnProLeuLeuGly--------------------- 14
               ||||||   ||||||         :::||||||||||||
Db          55 CCCGAGAGAGACCCGGCCATGCAGGAGCCGCTGCTGGGAGCCGAGGGCCCGGACTACGAC 114

Qy          15 ---------------ProProGlyGlySerAlaProArgGlyArg--------------- 24
                              ||||||               |||   |||
Db         115 ACCTTCCCCGAGAAGCCGCCCCCGTCGCCAGGGGACAGGGCGCGGGTCGGGACCCTGCAG 174

Qy          25 ------ArgValPheLeuAlaAlaPheAlaAlaAlaLeuGlyProLeuSerPheGlyPhe 42
                     |||||||||||||||   |||||||||   ||||||      |||||||||:::
Db         175 AACAAAAGGGTGTTCCTGGCCACCTTCGCCGCAGTGCTCGGCAATTTCAGCTTTGGGTAT 234

Qy          43 AlaLeuGlyTyrSerSerProAlaIleProSerLeuGlnArgAlaAlaProProAlaPro 62
               ||||||   |||:::||||||   ||||||:::|||:::|||:::      |||
Db         235 GCCCTGGTCTACACATCCCCTGTCATCCCAGCCCTGGAGCGCTCCTTGGATCCTGACCTG 294

Qy          63 ArgLeuAspAspAlaAlaAlaSerTrpPheGlyAlaValValThrLeuGlyAlaAlaAla 82
                  |||      :::   |||||||||||||||:::|||   ||||||||||||||||||
Db         295 CATCTGACCAAATCCCAGGCATCCTGGTTTGGGTCCGTGTTCACCCTGGGAGCAGCGGCC 354
```

```
Qy         83 GlyGlyValLeuGlyGlyTrpLeuValAspArgAlaGlyArgLysLeuSerLeuLeuLeu 102
              ||||||:::            |||   |||      |||||||||||||||::::::
Db        355 GGAGGCCTGAGTGCCATGATCCTCAACGACCTCCTGGGCCGGAAGCTGAGCATCATGTTC 414

Qy        103 CysSerValProPheValAlaGlyPheAlaValIleThrAlaAlaGlnAspValTrpMet 122
                 :::||||||      ||||||:::|||::::::      |||      :::||||||
Db        415 TCAGCTGTGCCGTCGGCGGCCGGCTATGCGCTCATGGCGGGTGCGCACGGCCTCTGGATG 474

Qy        123 LeuLeuGlyGlyArgLeuLeuThrGlyLeuAlaCysGlyValAlaSerLeuValAlaPro 142
              ||||||   ||||||   |||||||||   |||   |||:::   :::         |||
Db        475 CTGCTGCTCGGAAGGACGCTGACGGGCTTCGCCGGGGGGCTCACAGCTGCCTGCATCCCG 534

Qy        143 ValTyrIleSerGluIleAlaTyrProAlaValArgGlyLeuLeuGlySerCysValGln 162
              ||||||:::||||||||||||   |||   |||||||||   ||||||:::      |||
Db        535 GTGTACGTGTCTGAGATTGCTCCCCCAGGCGTTCGTGGGGCTCTGGGGGCCACACCCCAG 594

Qy        163 LeuMetValValValGlyIleLeuLeuAlaTyrLeuAlaGlyTrpValLeuGluTrpArg 182
              ||||||   |||   |||   |||      |||      |||   :::|||   ||||||
Db        595 CTCATGGCAGTGTTCGGATCCCTGTCCCTCTACGCCCTTGGCCTCCTGCTGCCGTGGCGC 654

Qy        183 TrpLeuAlaValLeuGlyCysValProProSerLeuMetLeuLeuLeuMetCysPheMet 202
              ||||||||||||   |||      |||      :::|||:::||||||:::   ||||||
Db        655 TGGCTGGCTGTGGCCGGGGAGGCGCCTGTGCTCATCATGATCCTGCTGCTCAGCTTCATG 714

Qy        203 ProGluThrProArgPheLeuLeuThrGlnHisArgArgGlnGluAlaMetAlaAlaLeu 222
              |||   :::|||||||||||||||::::::   |||   :::||||||:::   ||||||
Db        715 CCCAACTCGCCGCGCTTCCTGCTCTCTCGGGGCAGGGACGAAGAGGCCCTGCGGGCGCTG 774

Qy        223 ArgPheLeuTrpGlySerGluGlnGly------TrpGluAspProProIleGly------ 238
                 :::|||   |||::::::            ||||||         |||
Db        775 GCCTGGCTGCGTGGGACGGACGTCGATGTCCACTGGGAGTTCGAGCAGATCCAGGACAAC 834

Qy        239 ---AlaGluGlnSerPheHisLeu------AlaLeuLeuArgGlnProGlyIleTyrLys 255
                       ||||||      :::      |||      |||   |||   :::   :::
Db        835 GTCCGGAGACAGAGCAGCCGAGTATCGTGGGCTGAGGCACGGGCCCCACACGTGTGCCGG 894

Qy        256 ProPheIleIleGlyValSerLeuMetAlaPheGlnGlnLeuSerGlyValAsnAlaVal 275
              |||      :::   :::   :::         |||||||||:::|||:::      :::
Db        895 CCCATCACCGTGGCCTTGCTGATGCGCCTCCTGCAGCAGCTGACGGGCATCACGCCCATC 954

Qy        276 MetPheTyrAlaGluThrIlePheGluGluAlaLysPhe------------LysAspSer 291
              :::   |||   ::::::||||||:::                        ||||||
Db        955 CTGGTCTACCTGCAGTCCATCTTCGACAGCACCGCTGTCCTGCTGCCCCCCAAGGACGAC 1014

Qy        292 SerLeuAlaSerValValValGlyValIleGlnValLeuPheThrAlaValAlaAlaLeu 311
                       :::   :::||||||   :::::::::|||         :::|||||||||
Db       1015 ---------GCAGCCATCGTTGGGGCCGTGCGGCTCCTGTCCGTGCTGATCGCCGCCCTC 1065

Qy        312 IleMetAspArgAlaGlyArgArgLeuLeuLeuValLeuSerGlyValValMetValPhe 331
                 ||||||   |||||||||::::::||||||   :::|||      :::|||
Db       1066 ACCATGGACCTCGCAGGCCGCAAGGTGCTGCTCTTCGTCTCAGCGGCCATCATGTTTGCT 1125

Qy        332 SerThrSerAlaPheGlyAlaTyrPheLysLeuThrGlnGlyGlyProGlyAsnSerSer 351
              :::            |||   |||                  ||||||         |||
Db       1126 GCCAACCTGACTCTGGGGCTGTACATCCACTTT---------GGCCCCAGGCCTCTGAGC 1176

Qy        352 HisValAlaIleSerAlaProValSer------------------AlaGlnProValAsp 365
                    :::   :::         |||                  |||||||||:::
Db       1177 CCCAACAGCACTGCGGGCCTGGAAAGCGAGTCCTGGGGGACTTGGCGCAGCCCCTGGCA 1236

Qy        366 AlaSerValGly---LeuAlaTrpLeuAlaValGlySerMetCysLeuPheIleAlaGly 384
              |||      |||   |||      :::   :::   :::      |||||||||   |||
Db       1237 GCACCCGCTGGCTACCTCACCCTGGTGCCCCTGCTGGCCACCATGCTCTTCATCATGGGC 1296

Qy        385 PheAlaValGlyTrpGlyProIleProTrpLeuLeuMetSerGluIlePheProLeuHis 404
              :::|||||||||||||||||||||   ||||||||||||||||||:::   ||||||
Db       1297 TACGCCGTGGGCTGGGGTCCCATCACCTGGCTGCTCATGTCTGAGGTCCTGCCCCTGCGT 1356

Qy        405 ValLysGlyValAlaThrGlyIleCysValLeuThrAsnTrpLeuMetAlaPheLeuVal 424
                 :::|||||||||:::|||:::|||||||||   :::||||||   ||||||::::::
Db       1357 GCCCGTGGCGTGGCCTCAGGGCTCTGCGTGCTGGCCAGCTGGCTCACCGCCTTCGTCCTC 1416

Qy        425 ThrLysGluPheSerSerLeuMetGluValLeuArgProTyrGlyAlaPheTrpLeuAla 444
              ||||||   |||      :::   |||   |||||||||   |||||||||   |||
Db       1417 ACCAAGTCCTTCCTGCCAGTGGT-GAGCACCTTCGGCCTCCAGGTGCCTTTCTTCTTCTT 1475

Qy        445 -SerAlaPheCysIlePheSerValLeuPheThrLeuPheCysValProGluThrLysGl 464
               :::|||   |||:::   |||::::::||||||      ||||||||||||||||||||
Db       1476 CGCGGCCATCTGCTTGGTGAGCCTGGTGTTCACAGGCTGCTGTGTGCCCGAGACCAAGGG 1535

Qy        464 yLysThrLeuGluGlnIleThrAlaHisPheGlu---GlyArg 477
              |::::::||||||||||||   :::   |||      ||||||
Db       1536 ACGGTCCCTGGAGCAGATCGAGTCCTTCTTCCGCACGGGGAGA 1578

RESULT 12
US-10-424-599-132422
; Sequence 132422, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa Thomas J
;  APPLICANT:  Kovalic David K
;  APPLICANT:  Zhou Yihua
;  APPLICANT:  Cao Yongwei
;  TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53223)B
;  CURRENT APPLICATION NUMBER: US/10/424,599
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 132422
;   LENGTH: 2202
;   TYPE: DNA
;   ORGANISM: Glycine max
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(2202)
;   OTHER INFORMATION: unsure at all n locations
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT3847_90584C.1
US-10-424-599-132422

Alignment Scores:
Pred. No.:              1.2e-66        Length:          2202
Score:                  716.00         Matches:         164
Percent Similarity:     54.88%         Conservative:    89
Best Local Similarity:  35.57%         Mismatches:      174
Query Match:            29.14%         Indels:          34
DB:                     13             Gaps:            7

US-09-886-954A-1 (1-477) x US-10-424-599-132422 (1-2202)

Qy         26 ValPheLeuAlaAlaPheAlaAlaAlaLeuGlyProLeuSerPheGlyPheAlaLeuGly 45
              |||                     ||||||||||||:::   |||||||||      |||
Db        384 GTCCTCTTCTGCGTCCTCATCGTCGCCTTGGGTCCCATTCAATTCGGCTTCACGTGTGGG 443

Qy         46 TyrSerSerProAlaIleProSerLeuGlnArgAlaAlaProProAlaProArgLeuAsp 65
              ||||||||||||         ::::::   |||                     |||
Db        444 TATTCTTCTCCAACCCAAGGGGCTATAGTTCGCGATCTA------------AACCTCTCT 491

Qy         66 AspAlaAlaAlaSerTrpPheGlyAlaValValThrLeuGlyAlaAlaAlaGlyGlyVal 85
                 :::      |||:::||||||::::::      :::||||||      |||   :::
Db        492 ATTTCGGAGTTTTCTTTCTTTGGATCTTTGTCTAATGTGGGAGCAATGGTGGGAGCTATA 551

Qy         86 LeuGlyGlyTrpLeuValAspArgAlaGlyArgLysLeuSerLeuLeuLeuCysSerVal 105
                    |||   :::   :::      |||||||||   ||||||::::::   ::::::
Db        552 GCTAGTGGTCAGATAGCTGAATACATCGGGCGCAAAGGGTCATTGATGATTGCTGCGATC 611

Qy        106 ProPheValAlaGlyPheAlaValIleThrAlaAlaGlnAspValTrpMetLeuLeuGly 125
              |||   :::   |||:::      |||:::   |||:::|||         |||
Db        612 CCCAATATAATAGGGTGGCTTGCTATTTCTTTTGCCAAAGATTCCTCGTTTTTGTATATG 671

Qy        126 GlyArgLeuLeuThrGlyLeuAlaCysGlyValAlaSerLeuValAlaProValTyrIle 145
              ||||||||||||   |||         |||:::   |||   |||   ||||||||||||
Db        672 GGGAGGTTGTTGGAAGGTTTTGGCGTTGGGATTATCTCTTATGTGGTGCCTGTTTATATA 731
```

```
Qy        146 SerGluIleAlaTyrProAlaValArgGlyLeuLeuGlySerCysValGlnLeuMetVal 165
              :::|||||||||         :::||||||   |||||||||      ||||||   |||
Db        732 GCTGAGATTGCACCTCAAAACTTGAGAGGTGGCCTTGGGTCAGTGAACCAGCTCTCTGTT 791

Qy        166 ValValGlyIleLeuLeuAlaTyrLeuAlaGlyTrpValLeuGluTrpArgTrpLeuAla 185
                 :::||||||:::||||||||||||   |||      :::   ||||||   ||||||
Db        792 ACAATTGGCATTATGCTGGCTTATCTGTTGGGTCTTTTTGTCAACTGGAGAGTCCTTGCA 851

Qy        186 ValLeuGlyCysValProProSerLeuMetLeuLeuLeuMetCysPheMetProGluThr 205
              :::||||||   :::|||   ::::::::::::      :::   |||:::||||||:::
Db        852 ATTCTAGGAATTTTGCCTTGTACAGTATTAATACCTGGATTATTTTCATACCTGAATCC 911

Qy        206 ProArgPheLeuLeuThrGlnHisArgArgGlnGluAlaMetAlaAlaLeuArgPheLeu 225
              ||||||:::|||                     |||         :::|||:::   |||
Db        912 CCCAGATGGTTGGCCAAGATGGGGATGATAGATGAGTTTGAGACTTCTTTGCAAGTGTTA 971

Qy        226 TrpGlySerGluGlnGlyTrpGluAspProProIle------------------------ 237
                 |||   :::            |||      :::
Db        972 CGAGGATTTGAC---------ACTGATATATCTGTTGAAGTACATGAAATTAAGAGATCT 1022

Qy        238 ------------GlyAlaGluGlnSerPheHisLeuAlaLeuLeuArgGlnProGlyIle 253
                          |||         :::         |||   |||::::::
Db       1023 GTGGCTTCAACGGGAAAAAGAGCTGCAATCCGATTTGCAGATCTCAAGAGGAAAAGATAT 1082

Qy        254 TyrLysProPheIleIleGlyValSerLeuMetAlaPheGlnGlnLeuSerGlyValAsn 273
              :::   |||   ::::::|||:::   |||:::      |||||||||||||||:::|||
Db       1083 TGGTTCCCGTTAATGGTTGGTATTGGATTACTTGTTCTTCAGCAATTATCTGGTATCAAT 1142

Qy        274 AlaValMetPheTyrAlaGluThrIlePheGluGluAlaLysPheLysAspSerSerLeu 293
                 ::::::||||||:::      |||||||||      |||            |||
Db       1143 GGAATTTTGTTCTATTCAACTACCATCTTTGCAAATGCAGGAATTTCATCCAGCGAAGCT 1202

Qy        294 AlaSerValValValGlyValIleGlnValLeuPheThrAlaValAlaAlaLeuIleMet 313
              |||:::|||   :::|||   :::||||||:::   |||   ::::::      ::::::
Db       1203 GCTACAGTTGGACTTGGAGCCGTTCAGGTCATAGCAACTGGAATTTCCACATGGTTGGTG 1262

Qy        314 AspArgAlaGlyArgArgLeuLeuLeuValLeuSerGlyValValMetValPheSerThr 333
              |||::::::||||||||||||||||||::::::|||      ||||||      |||
Db       1263 GACAAAAGTGGCCGGAGGCTGCTTCTAATAATATCCTCATCTGTAATGACAGTTAGCCTT 1322

Qy        334 SerAlaPheGlyAlaTyrPheLysLeuThrGlnGlyGlyProGlyAsnSerSerHisVal 353
                                |||   |||   :::|||               ||||||:::
Db       1323 CTCATTGTTTCTATAGCATTTTATCTG---GAGGGGGTTGTATCAGAGGATTCACATTTA 1379

Qy        354 AlaIleSerAlaProValSerAlaGlnProValAspAlaSerValGlyLeuAlaTrpLeu 373
                                |||         :::   :::   |||||||||      :::
Db       1380 ---------------TTCAGCATTTTGGGAATAGTTTCTATTGTTGGACTCGTGGCTATG 1424

Qy        374 AlaValGlySerMetCysLeuPheIleAlaGlyPheAlaValGlyTrpGlyProIlePro 393
                 :::                        ||||||::::::|||   ||||||||||||
Db       1425 GTGATT------------------------GGGTTCTCTCTAGGTCTGGGACCCATCCCT 1460

Qy        394 TrpLeuLeuMetSerGluIlePheProLeuHisValLysGlyValAlaThrGlyIleCys 413
              ||||||:::||||||||||||   |||:::::::::||||||:::|||      |||
Db       1461 TGGCTTATAATGTCTGAGATACTTCCAGTGAATATAAAGGGCCTTGCTGGCAGCATAGCG 1520

Qy        414 ValLeuThrAsnTrpLeuMetAlaPheLeuValThrLysGluPheSerSerLeuMetGlu 433
                 :::   |||||||||:::::::::   :::|||         :::   |||:::
Db       1521 ACAATGGGAAATTGGCTGATTTCGTGGGGGATCACGATGACTGCTAACTTGCTTTTGAAT 1580

Qy        434 ValLeuArgProTyrGlyAlaPheTrpLeuAlaSerAlaPheCysIlePheSerValLeu 453
                             |||   |||   :::   :::            |||::::::
Db       1581 ---TGGAGCAGTGGAGGGACATTTACAATCTACACAGTCGTAGCTGCCTTTACTATAGCT 1637

Qy        454 PheThrLeuPheCysValProGluThrLysGlyLysThrLeuGluGlnIleThrAlaHis 473
              |||            ||||||||||||||||||:::|||||||||:::|||
Db       1638 TTTATAGCAATGTGGGTTCCTGAGACCAAGGGAAGAACATTGGAAGAAATTCAGTTTTCC 1697

Qy        474 Phe 474
              |||
Db       1698 TTC 1700

RESULT 13
```

```
US-10-425-114-26596
; Sequence 26596, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
;  APPLICANT: Liu, Jingdong
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Screen, Steven E
;  APPLICANT:  Tabaska, Jack E
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53313)B
;  CURRENT APPLICATION NUMBER: US/10/425,114
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 26596
;   LENGTH: 1797
;   TYPE: DNA
;   ORGANISM: Oryza sativa
;   FEATURE:
;   OTHER INFORMATION: Clone ID: LIB4371-017-F11_FLI
US-10-425-114-26596

Alignment Scores:
Pred. No.:               2.9e-65        Length:         1797
Score:                   702.00         Matches:        170
Percent Similarity:      53.56%         Conservative:   78
Best Local Similarity:   36.72%         Mismatches:     176
Query Match:             28.57%         Indels:         40
DB:                      13             Gaps:           6

US-09-886-954A-1 (1-477) x US-10-425-114-26596 (1-1797)

Qy         27 PheLeuAlaAlaPheAlaAlaAlaLeuGlyProLeuSerPheGlyPheAlaLeuGlyTyr 46
              ||||||               ||||||||||||:::   |||||||||      |||:::
Db        284 TTCCTCTGCACGCTCATCGTCGCGCTCGGCCCCATCCAGTTCGGATTCACGGGGGGCTTC 343

Qy         47 SerSerProAlaIleProSerLeuGlnArgAlaAlaProProAlaProArg-LeuAspAs 66
              |||||||||               |||   |||      |||   |||||| |||
Db        344 TCCTCCCCGACG------------CAGGACGCCATCAT-CCGAGACCTCGACCTCACCCT 390

Qy         66 pAlaAlaAlaSerTrpPheGlyAlaValValThrLeuGlyAlaAlaAlaGlyGlyValLe 86
               :::      |||   ||||||::::::      :::||||||      |||   :::
Db        391 CTCCGAGTTCTCGGTGTTCGGATCGCTGTCCAACGTCGGCGCCATGGTTGGGGCGATTGC 450

Qy         86 uGlyGlyTrpLeuValAspArgAlaGlyArgLysLeuSerLeuLeuLeuCysSerValPr 106
                  |||   :::   :::      |||||||||   ||||||::::::   ::::::||
Db        451 CAGTGGTCAGATGGCCGAGTACATTGGGCGCAAAGGGTCATTGATGATTGCTGCAATTCC 510

Qy        106 oPheValAlaGlyPheAlaValIleThrAlaAlaGlnAspValTrpMetLeuLeuGlyGl 126
              |   :::   |||:::      |||:::   |||:::|||         |||      ||
Db        511 AAACATCATTGGTTGGCTTGCCATCTCCTTTGCAAAAGACTCATCGTTTCTTTATATGGG 570

Qy        126 yArgLeuLeuThrGlyLeuAlaCysGlyValAlaSerLeuValAlaProValTyrIleSe 146
              ||||||||||   |||         ||||||   |||         ||||||||||||::
Db        571 ACGATTGCTCGAGGGGTTTGGTGTTGGTGTCATCTCTTATACGGTGCCAGTTTACATAGC 630

Qy        146 rGluIleAlaTyrProAlaValArgGlyLeuLeuGlySerCysValGlnLeuMetValVa 166
              :||||||:::         :::||||||   |||||||||      ||||||   |||
Db        631 AGAAATATCACCTCAAAACATGAGAGGTGCTCTTGGCTCAGTGAATCAGTTATCTGTAAC 690

Qy        166 lValGlyIleLeuLeuAlaTyrLeuAlaGlyTrpValLeuGluTrpArgTrpLeuAlaVa 186
               ||||||||||||||||||||||||   |||      :::   ||||||   ||||||||
Db        691 CGTTGGTATATTGTTGGCATATTTGCTCGGCATGTTTGTTCCTTGGAGGCTTCTTGCTGT 750

Qy        186 lLeuGlyCysValProProSerLeuMetLeuLeuLeuMetCysPheMetProGluThrPr 206
              |:::|||   :::|||   ::::::::::::      :::   |||:::||||||:::||
Db        751 AATAGGAATCTTGCCTTGCACTGTGTTGATACCTGGCCTATTCTTCATTCCAGAATCCCC 810

Qy        206 oArgPheLeuLeuThrGlnHisArgArgGlnGluAlaMetAlaAlaLeuArgPheLeuTr 226
              ||||:::|||         :::         :::         :::|||:::   |||
Db        811 AAGATGGTTGGCAAAGATGAACATGATGGATGATTTTGAGACTTCTTTACAAGTTCTGAG 870
```

```
Qy         226 pGlySerGluGlnGlyTrpGluAspProProIleGlyAlaGlu----------------- 240
                |||   |||                     |||   ||||||
Db         871 GGGATTTGAG---------------ACTGACATCAGCGCGGAAGTGAATGATATAAAGAG 915

Qy         241 -------------------------GlnSerPheHisLeuAlaLeuLeuArgGlnProGl 252
                                           :::               |||   |||
Db         916 AGCAGTAGCGTCAGCAAACAAAAGGACAACGATCCGTTTTCAAGAATTAAACCAGAAGAA 975

Qy         252 yIleTyrLysProPheIleIleGlyValSerLeuMetAlaPheGlnGlnLeuSerGlyVa 272
                         |||   |||:::|||:::   |||:::      |||||||||||||||::
Db         976 ATACCGCACACCCCTAATACTAGGAATTGGCCTACTTGTACTGCAACAGCTAAGTGGAAT 1035

Qy         272 lAsnAlaValMetPheTyrAlaGluThrIlePheGluGluAlaLysPheLysAspSerSe 292
               :|||   ::::::|||||||||   :::||||||:::   |||         :::|||
Db        1036 CAATGGAATATTGTTTTATGCAGGTAGCATCTTCAAAGCAGCAGGTCTCACAAACAGTGA 1095

Qy         292 rLeuAlaSerValValValGlyValIleGlnValLeuPheThrAlaValAlaAlaLeuIl 312
                ||||||:::      :::|||   ||||||||||||   |||   |||         ::
Db        1096 CTTGGCTACATGTGCACTTGGTGCTATCCAGGTTCTTGCTACAGGAGTTACAACCTGGTT 1155

Qy         312 eMetAspArgAlaGlyArgArgLeuLeuLeuValLeuSerGlyValValMetValPheSe 332
               ::::||||||||||||||||||:::||||||::::::|||         |||      ||
Db        1156 ATTAGACAGAGCTGGCCGACGGATCCTCCTTATCATCTCTTCTGCTGGGATGACTCTAAG 1215

Qy         332 rThrSerAlaPheGlyAlaTyrPheLysLeuThrGlnGlyGlyProGlyAsnSerSerHi 352
               |      |||            |||   |||   :::                  |||||
Db        1216 CCTCCTTGCAGTTGCTGTTGTATTTTTCCTC---AAGGATAGCATTTCACAAGATTCTCA 1272

Qy         352 sValAlaIleSerAlaProValSerAlaGlnProValAspAlaSerValGlyLeuAlaTr 372
               |:::      :::                                       |||:::
Db        1273 CATGTACTACACC---------------------------------------TTAAGTAT 1293

Qy         372 pLeuAlaValGlySerMetCysLeuPheIleAlaGlyPheAlaValGlyTrpGlyProIl 392
                :::::::::   ::::::      |||:::      |||:::   |||   |||   ||
Db        1294 GATCTCCTTGGTTGCTCTTGTGGCTTTTGTAATCGCCTTCTCCTTCGGTATGGGTGCCAT 1353

Qy         392 eProTrpLeuLeuMetSerGluIlePheProLeuHisValLysGlyValAlaThrGlyIl 412
               |||||||::::::||||||||||||   |||:::   :::|||   :::|||
Db        1354 TCCATGGATCATAATGTCAGAGATCCTCCCGGTTAGTATCAAGAGTCTCGCAGGAAGCTT 1413

Qy         412 eCysValLeuThrAsnTrpLeuMetAlaPheLeuValThrLysGluPheSerSerLeuMe 432
                      |||   |||||||||   :::|||   :::|||         ::::::|||||
Db        1414 TGCGACGCTCGCCAACTGGCTTACATCCTTTGGAATAACAATGACA---GCAAACTTGAT 1470

Qy         432 tGluValLeuArgProTyrGlyAlaPheTrpLeuAlaSerAlaPheCysIlePheSerVa 452
               |                  |||   |||                        |||:::::
Db        1471 GCTTAGCTGGAGTGCTGGAGGGACCTTTGTGTCCTACATGGTCGTGAGTGCTTTCACCCT 1530

Qy         452 lLeuPheThrLeuPheCysValProGluThrLysGlyLysThrLeuGluGlnIleThrAl 472
               ::::|||   :::      ||||||||||||||||||:::|||||||||:::|||
Db        1531 CGTGTTCGTCATCCTTTGGGTGCCAGAGACAAAAGGAAGAACTCTCGAAGAGATACAATG 1590

Qy         472 aHisPhe 474
                   |||
Db        1591 GTCCTTC 1597

RESULT 14
US-10-437-963-61061
; Sequence 61061, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  APPLICANT:  Wu, Wei
;  APPLICANT:  Boukharov, Andrey A.
;  APPLICANT:  Barbazuk, Brad
;  APPLICANT:  Li, Ping
;  TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53221)B
```

```
;  CURRENT APPLICATION NUMBER: US/10/437,963
;  CURRENT FILING DATE:  2003-05-14
;  NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 61061
;   LENGTH: 2003
;   TYPE: DNA
;   ORGANISM: Oryza sativa
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_62530C.1
US-10-437-963-61061

Alignment Scores:
Pred. No.:              3.38e-65      Length:        2003
Score:                  702.00        Matches:       170
Percent Similarity:     53.56%        Conservative:  78
Best Local Similarity:  36.72%        Mismatches:    176
Query Match:            28.57%        Indels:        40
DB:                     17            Gaps:          6

US-09-886-954A-1 (1-477) x US-10-437-963-61061 (1-2003)

Qy          27 PheLeuAlaAlaPheAlaAlaAlaLeuGlyProLeuSerPheGlyPheAlaLeuGlyTyr 46
               ||||||               ||||||||||||:::   |||||||||      |||:::
Db         346 TTCCTCTGCACGCTCATCGTCGCGCTCGGCCCCATCCAGTTCGGATTCACGGGGGGCTTC 405

Qy          47 SerSerProAlaIleProSerLeuGlnArgAlaAlaProProAlaProArg-LeuAspAs 66
               |||||||||               |||   |||      |||   |||||| |||
Db         406 TCCTCCCCGACG------------CAGGACGCCATCAT-CCGAGACCTCGACCTCACCCT 452

Qy          66 pAlaAlaAlaSerTrpPheGlyAlaValValThrLeuGlyAlaAlaAlaGlyGlyValLe 86
                :::      |||   ||||||::::::      :::||||||      |||   :::
Db         453 CTCCGAGTTCTCGGTGTTCGGATCGCTGTCCAACGTCGGCGCCATGGTTGGGGCGATTGC 512

Qy          86 uGlyGlyTrpLeuValAspArgAlaGlyArgLysLeuSerLeuLeuLeuCysSerValPr 106
                   |||   :::   :::      |||||||||   ||||||::::::   ::::::||
Db         513 CAGTGGTCAGATGGCCGAGTACATTGGGCGCAAAGGGTCATTGATGATTGCTGCAATTCC 572

Qy         106 oPheValAlaGlyPheAlaValIleThrAlaAlaGlnAspValTrpMetLeuLeuGlyGl 126
               |   :::   |||:::      |||:::   |||:::|||         |||      ||
Db         573 AAACATCATTGGTTGGCTTGCCATCTCCTTTGCAAAAGACTCATCGTTTCTTTATATGGG 632

Qy         126 yArgLeuLeuThrGlyLeuAlaCysGlyValAlaSerLeuValAlaProValTyrIleSe 146
               ||||||||||   |||         ||||||   |||         ||||||||||||::
Db         633 ACGATTGCTCGAGGGGTTTGGTGTTGGTGTCATCTCTTATACGGTGCCAGTTTACATAGC 692

Qy         146 rGluIleAlaTyrProAlaValArgGlyLeuLeuGlySerCysValGlnLeuMetValVa 166
               :||||||:::         :::||||||   |||||||||      ||||||   |||
Db         693 AGAAATATCACCTCAAAACATGAGAGGTGCTCTTGGCTCAGTGAATCAGTTATCTGTAAC 752

Qy         166 lValGlyIleLeuLeuAlaTyrLeuAlaGlyTrpValLeuGluTrpArgTrpLeuAlaVa 186
                ||||||||||||||||||||||||   |||      :::   ||||||   ||||||||
Db         753 CGTTGGTATATTGTTGGCATATTTGCTCGGCATGTTTGTTCCTTGGAGGCTTCTTGCTGT 812

Qy         186 lLeuGlyCysValProProSerLeuMetLeuLeuLeuMetCysPheMetProGluThrPr 206
               |:::|||   :::|||   ::::::::::::      :::   |||:::||||||:::||
Db         813 AATAGGAATCTTGCCTTGCACTGTGTTGATACCTGGCCTATTCTTCATTCCAGAATCCCC 872

Qy         206 oArgPheLeuLeuThrGlnHisArgArgGlnGluAlaMetAlaAlaLeuArgPheLeuTr 226
               ||||:::|||         :::         :::         :::|||:::   |||
Db         873 AAGATGGTTGGCAAAGATGAACATGATGGATGATTTTGAGACTTCTTTACAAGTTCTGAG 932

Qy         226 pGlySerGluGlnGlyTrpGluAspProProIleGlyAlaGlu----------------- 240
                |||   |||                     |||   ||||||
Db         933 GGGATTTGAG---------------ACTGACATCAGCGCGGAAGTGAATGATATAAAGAG 977

Qy         241 -------------------------GlnSerPheHisLeuAlaLeuLeuArgGlnProGl 252
                                           :::               |||   |||
Db         978 AGCAGTAGCGTCAGCAAACAAAAGGACAACGATCCGTTTTCAAGAATTAAACCAGAAGAA 1037

Qy         252 yIleTyrLysProPheIleIleGlyValSerLeuMetAlaPheGlnGlnLeuSerGlyVa 272
                         |||   |||:::|||:::   |||:::      |||||||||||||||::
Db        1038 ATACCGCACACCCCTAATACTAGGAATTGGCCTACTTGTACTGCAACAGCTAAGTGGAAT 1097
```

```
Qy         272 lAsnAlaValMetPheTyrAlaGluThrIlePheGluGluAlaLysPheLysAspSerSe 292
               :|||   ::::::|||||||||   :::||||||:::   |||         :::|||
Db        1098 CAATGGAATATTGTTTTATGCAGGTAGCATCTTCAAAGCAGCAGGTCTCACAAACAGTGA 1157

Qy         292 rLeuAlaSerValValValGlyValIleGlnValLeuPheThrAlaValAlaAlaLeuIl 312
                ||||||:::      :::|||   ||||||||||||   |||   |||         ::
Db        1158 CTTGGCTACATGTGCACTTGGTGCTATCCAGGTTCTTGCTACAGGAGTTACAACCTGGTT 1217

Qy         312 eMetAspArgAlaGlyArgArgLeuLeuLeuValLeuSerGlyValValMetValPheSe 332
               ::::||||||||||||||||||:::||||||::::::|||         |||      ||
Db        1218 ATTAGACAGAGCTGGCCGACGGATCCTCCTTATCATCTCTTCTGCTGGGATGACTCTAAG 1277

Qy         332 rThrSerAlaPheGlyAlaTyrPheLysLeuThrGlnGlyGlyProGlyAsnSerSerHi 352
               |      |||            |||   |||   :::                  |||||
Db        1278 CCTCCTTGCAGTTGCTGTTGTATTTTTCCTC---AAGGATAGCATTTCACAAGATTCTCA 1334

Qy         352 sValAlaIleSerAlaProValSerAlaGlnProValAspAlaSerValGlyLeuAlaTr 372
               |:::      :::                                       |||:::
Db        1335 CATGTACTACACC---------------------------------------TTAAGTAT 1355

Qy         372 pLeuAlaValGlySerMetCysLeuPheIleAlaGlyPheAlaValGlyTrpGlyProIl 392
                :::::::::   ::::::      |||:::      |||:::   |||   |||   ||
Db        1356 GATCTCCTTGGTTGCTCTTGTGGCTTTTGTAATCGCCTTCTCCTTCGGTATGGGTGCCAT 1415

Qy         392 eProTrpLeuLeuMetSerGluIlePheProLeuHisValLysGlyValAlaThrGlyIl 412
               |||||||::::::||||||||||||   |||:::   :::|||   :::|||
Db        1416 TCCATGGATCATAATGTCAGAGATCCTCCCGGTTAGTATCAAGAGTCTCGCAGGAAGCTT 1475

Qy         412 eCysValLeuThrAsnTrpLeuMetAlaPheLeuValThrLysGluPheSerSerLeuMe 432
                      |||   |||||||||   :::|||   :::|||         ::::::|||||
Db        1476 TGCGACGCTCGCCAACTGGCTTACATCCTTTGGAATAACAATGACA---GCAAACTTGAT 1532

Qy         432 tGluValLeuArgProTyrGlyAlaPheTrpLeuAlaSerAlaPheCysIlePheSerVa 452
               |                  |||   |||                        |||:::::
Db        1533 GCTTAGCTGGAGTGCTGGAGGGACCTTTGTGTCCTACATGGTCGTGAGTGCTTTCACCCT 1592

Qy         452 lLeuPheThrLeuPheCysValProGluThrLysGlyLysThrLeuGluGlnIleThrAl 472
               ::::|||   :::      ||||||||||||||||||:::|||||||||:::|||
Db        1593 CGTGTTCGTCATCCTTTGGGTGCCAGAGACAAAAGGAAGAACTCTCGAAGAGATACAATG 1652

Qy         472 aHisPhe 474
                   |||
Db        1653 GTCCTTC 1659
```

```
RESULT 15
US-10-425-114-24968
; Sequence 24968, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
;  APPLICANT: Liu, Jingdong
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Screen, Steven E
;  APPLICANT:  Tabaska, Jack E
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53313)B
;  CURRENT APPLICATION NUMBER: US/10/425,114
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 24968
;   LENGTH: 2049
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   OTHER INFORMATION: Clone ID: LIB3689-227-G11_FLI
US-10-425-114-24968

Alignment Scores:
Pred. No.:              3.96e-65      Length:         2049
Score:                  701.50        Matches:        173
Percent Similarity:     51.19%        Conservative:   86
Best Local Similarity:  34.19%        Mismatches:     182
Query Match:            28.55%        Indels:         65
DB:                     13            Gaps:           8

US-09-886-954A-1 (1-477) x US-10-425-114-24968 (1-2049)
```

```
Qy          10 GlnProLeuLeuGly---------------------ProProGlyGlySerAlaProArg 22
               :::||||||:::                        ||||||   |||
Db         390 AAGCCGCTCATCAACACCGGGAGCTGGTACCGCATGCCGCCGGCGGGTGGCGTGATGGGC 449

Qy          23 GlyArg------------------ArgValPheLeuAlaAlaPheAla------------ 32
                  |||                  |||:::      :::||||||:::
Db         450 TCGCGGCAATCCAGCCTCATGGAGCGCTTGGGCTCCTCTGCGTTCTCTCTCCGCGACGTC 509

Qy          33 ------------------------------AlaAlaLeuGlyProLeuSerPheGlyPhe 42
                                                ||||||||||||:::   |||||||||
Db         510 GCTATCTCGGCCACGCTCTGCACGCTTATTGTCGCGCTAGGTCCCATCCAGTTCGGTTTC 569

Qy          43 AlaLeuGlyTyrSerSerProAlaIleProSerLeuGlnArgAlaAlaProProAlaPro 62
                     |||||||||||||||               |||   |||
Db         570 ACATGCGGCTACTCCTCGCCCACG------------CAGGACGCCATCATTGCTGATCTC 617

Qy          63 ArgLeuAspAspAlaAlaAlaSerTrpPheGlyAlaValValThrLeuGlyAlaAlaAla 82
                  |||      :::      |||   ||||||::::::      :::||||||
Db         618 GGCCTCTCCCTCTCTGAGTTCTCCCTCTTCGGTTCATTATCTAATGTAGGGGCGATGGTA 677

Qy          83 GlyGlyValLeuGlyGlyTrpLeuValAspArgAlaGlyArgLysLeuSerLeuLeuLeu 102
               |||   :::      |||   |||   :::      |||||||||   ||||||::::::
Db         678 GGCGCCATCTCCAGTGGGCAACTTGCAGAGTATATCGGCCGCAAGGGGTCTCTCATGATC 737

Qy         103 CysSerValProPheValAlaGlyPheAlaValIleThrAlaAlaGlnAspValTrpMet 122
                  ::::::|||   :::   |||:::      |||:::   |||:::|||
Db         738 GCTGCGATTCCAAACATAATTGGGTGGCTCGCGATATCATTCGCAAAAGATTCCTCTTTC 797

Qy         123 LeuLeuGlyGlyArgLeuLeuThrGlyLeuAlaCysGlyValAlaSerLeuValAlaPro 142
               |||      ||||||||||||   |||         ||||||   |||         |||
Db         798 TTGTTTATGGGTCGGCTGCTAGAAGGATTTGGAGTCGGTGTAATATCGTATACAGTACCG 857

Qy         143 ValTyrIleSerGluIleAlaTyrProAlaValArgGlyLeuLeuGlySerCysValGln 162
               |||||||||:::|||||||||            ||||||   |||||||||      |||
Db         858 GTTTATATTGCAGAAATCGCTCCTCAGGATCAGAGGGGAGCTCTTGGTTCTGTCAATCAG 917

Qy         163 LeuMetValValValGlyIleLeuLeuAlaTyrLeuAlaGlyTrpValLeuGluTrpArg 182
               |||   |||   :::|||||||||||||||||||||   |||      :::   ||||||
Db         918 CTCTCCGTCACGATTGGTATATTGCTTGCCTACTTGTTTGGCATGTTTGTTCCCTGGAGA 977

Qy         183 TrpLeuAlaValLeuGlyCysValProProSerLeuMetLeuLeuLeuMetCysPheMet 202
                  |||||||||||||||   :::|||   |||:::::::::      :::   |||:::
Db         978 ATTCTTGCTGTTCTAGGCATTTTACCTTGTTCAATCCTGATTCCTGGACTGTTCTTTGTG 1037

Qy         203 ProGluThrProArgPheLeuLeuThrGlnHisArgArgGlnGluAlaMetAlaAlaLeu 222
               ||||||:::||||||:::|||            :::   ::::::         :::|||
Db        1038 CCTGAATCCCCAAGGTGGTTGGCAAAAATGGGGAAGATGGAGGATTTTGAATATTCATTG 1097

Qy         223 ArgPheLeuTrpGlySerGluGlnGlyTrpGluAspProProIleGlyAlaGlu------ 240
               :::   |||   |||   :::                     |||   ||||||
Db        1098 CAAGTTCTGAGAGGATTTCAG---------------ACAGATATCACAGCAGAAGTAAAT 1142

Qy         241 ------------------------------------GlnSerPheHisLeuAlaLeuLeu 248
                                                      :::         |||   :::
Db        1143 GAAATAAAGAGATCATTAGCATCATCGAGGAGGAGGACAACCATAAGGTTCGCTGATATC 1202

Qy         249 ArgGlnProGlyIleTyrLysProPheIleIleGlyValSerLeuMetAlaPheGlnGln 268
               :::|||               |||   :::||||||:::   |||:::      ||||||
Db        1203 AAACAGAAGAGATACAGTGTTCCCCTTGTGATAGGAATCGGTCTCCTTGTCCTGCAGCAG 1262

Qy         269 LeuSerGlyValAsnAlaValMetPheTyrAlaGluThrIlePheGluGluAlaLysPhe 288
               |||||||||||||||   ::::::|||||||||   :::||||||:::   |||
Db        1263 CTAAGTGGTGTCAATGGCATTCTATTTTATGCTGCGAGCATCTTCAAAGCTGCTGGTATT 1322

Qy         289 LysAspSerSerLeuAlaSerValValValGlyValIleGlnValLeuPheThrAlaVal 308
                  :::|||:::||||||:::      :::|||   :::||||||:::   |||   |||
Db        1323 ACAAACAGTAATCTAGCAACATTTGGTTTAGGGGCTGTTCAGGTGATTGCTACTGGAGTG 1382
```

```
Qy          309 AlaAlaLeuIleMetAspArgAlaGlyArgArgLeuLeuLeuValLeuSerGlyValVal 328
                         :::   |||:::|||||||||||||||||||||::::::|||
Db         1383 ACAACCTGGTTGACTGACAAAGCTGGTCGAAGGCTTCTTCTCATTATTTCCACCACAGGA 1442

Qy          329 MetValPheSerThrSerAlaPheGlyAlaTyrPheLysLeuThrGlnGlyGlyProGly 348
                ||||||   :::                     |||   :::
Db         1443 ATGGTCATTACTCTTGTTATTGTTTCTGTGTCATTTTTTGTGAAGGACAACATAGCTGCT 1502

Qy          349 AsnSerSerHisValAlaIleSerAlaProValSerAlaGlnProValAspAlaSerVal 368
                   |||         ::::::
Db         1503 GGTTCGCACTTATACTCTGTA--------------------------------------- 1523

Qy          369 GlyLeuAlaTrpLeuAlaValGlySerMetCysLeuPheIleAlaGlyPheAlaValGly 388
                   ::::::   |||::::::      :::      |||:::      |||::::::|||
Db         1524 ---ATGAGTATGCTTTCACTGGCTGGACTTGTGGCATTTGTGATTGCATTTTCTCTTGGC 1580

Qy          389 TrpGlyProIleProTrpLeuLeuMetSerGluIlePheProLeuHisValLysGlyVal 408
                   |||   |||||||||::::::||||||||||||   |||:::::::::|||   :::
Db         1581 TTGGGAGCGATTCCGTGGATCATTATGTCTGAGATCCTTCCTGTTAACATCAAGAGCCTT 1640

Qy          409 AlaThrGlyIleCysValLeuThrAsnTrpLeuMetAlaPheLeuValThrLysGluPhe 428
                |||      :::      |||   |||||||||   |||:::   :::|||
Db         1641 GCTGGAAGTGTTGCGACCCTGGCGAACTGGCTGACAGCATGGGCCATTACAATGACG--- 1697

Qy          429 SerSerLeuMetGluValLeuArgProTyrGlyAlaPheTrpLeuAlaSerAlaPheCys 448
                :::|||||||||                  |||   |||   :::   :::
Db         1698 GCAAGCCTGATGTTGAACTGGAGCAGTGGAGGAACATTTGCTATCTACGCCGTCGTGTCT 1757

Qy          449 IlePheSerValLeuPheThrLeuPheCysValProGluThrLysGlyLysThrLeuGlu 468
                      :::::::::|||            ||||||||||||||||||:::|||||||||
Db         1758 ACCATGGCCCTCATTTCGTGTGCTTGTGGGTGCCTGAGACCAAGGGAAGAACGCTAGAG 1817

Qy          469 GlnIleThrAlaHisPhe 474
                :::|||         |||
Db         1818 GAAATCGCCTTCTCATTC 1835

Search completed: September 28, 2004, 08:02:21
Job time : 4267 secs
```

```
                    GenCore version 5.1.6
           Copyright (c) 1993 - 2004  Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on:         September 28, 2004, 01:47:14 ; Search time 2846 Seconds
                                          (without alignments)
                                          5005.011 Million cell updates/sec

Title:          US-09-886-954A-1
Perfect score:  2457
Sequence:       1 MTPEDPEETQPLLGPPGGSA..........CVPETKGKTLEQITAHFEGR 477

Scoring table:  BLOSUM62
                Xgapop 10.0 , Xgapext  0.5
                Ygapop 10.0 , Ygapext  0.5
                Fgapop  6.0 , Fgapext  7.0
                Delop   6.0 , Delext   7.0

Searched:       27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters:      55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US09886954/runat_27092004_170632_13400/app_query.fasta_1.647
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09886954_@CGN_1_1_5180_@runat_27092004_170632_13400 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :      EST:*
                1:  em_estba:*
                2:  em_esthum:*
                3:  em_estin:*
                4:  em_estmu:*
                5:  em_estov:*
                6:  em_estpl:*
                7:  em_estro:*
                8:  em_htc:*
                9:  gb_est1:*
                10:  gb_est2:*
                11:  gb_htc:*
                12:  gb_est3:*
                13:  gb_est4:*
                14:  gb_est5:*
                15:  em_estfun:*
                16:  em_estom:*
                17:  em_gss_hum:*
                18:  em_gss_inv:*
                19:  em_gss_pln:*
                20:  em_gss_vrt:*
                21:  em_gss_fun:*
                22:  em_gss_mam:*
                23:  em_gss_mus:*
                24:  em_gss_pro:*
                25:  em_gss_rod:*
                26:  em_gss_phg:*
                27:  em_gss_vrl:*
                28:  gb_gss1:*
                29:  gb_gss2:*

       Pred. No. is the number of results predicted by chance to have a
       score greater than or equal to the score of the result being printed,
       and is derived by analysis of the total score distribution.

                                  SUMMARIES
                         %
Result                 Query
  No.     Score       Match Length DB  ID                           Description
 -------------------------------------------------------------------------------
     1     2159        87.9   2101  11  AK081806                AK081806 Mus muscu
     2   2123.5        86.4   1954  11  AF289587                AF289587 Homo sapi
     3     1713        69.7   1008  29  AY414180                AY414180 Homo sapi
     4   1490.5        60.7   1201  13  BX395379                BX395379 BX395379
     5   1485.5        60.5   1000  29  AY414182                AY414182 Mus muscu
     6     1379        56.1   1085  12  BM545247                BM545247 AGENCOURT
     7     1340        54.5    932  12  BG479842                BG479842 602527373
     8     1293        52.6    830  12  BG700749                BG700749 602681616
     9   1269.5        51.7    991  12  BI334832                BI334832 602998959
    10     1269        51.6   1004  29  AY414181                AY414181 Pan trogl
    11     1267        51.6    875  14  CD557801                CD557801 AGENCOURT
    12   1241.5        50.5    880  12  BM044230                BM044230 603621486
    13     1238        50.4    828  12  BG478000                BG478000 602522606
    14     1233        50.2    777  12  BG717034                BG717034 602689093
    15   1230.5        50.1    889  12  BI757409                BI757409 603029344
    16   1222.5        49.8    879  12  BG771736                BG771736 602720404
    17   1193.5        48.6    838  12  BG281777                BG281777 602402292
    18   1183.5        48.2    801  12  BI114765                BI114765 602861314
    19   1120.5        45.6    894  13  BQ919165                BQ919165 AGENCOURT
    20     1119        45.5    899  12  BI916408                BI916408 603178364
    21     1116        45.4    854  12  BI828656                BI828656 603078566
    22     1076        43.8    813  12  BI831908                BI831908 603078776
    23   1074.5        43.7   1002  12  BI828949                BI828949 603074819
    24     1059        43.1    833  12  BI826390                BI826390 603076257
c   25     1055        42.9    700  13  BQ109572                BQ109572 imageqc_7
    26   1036.5        42.2    934  10  BE910478                BE910478 601501005
    27   1030.5        41.9    773  13  BQ603775                BQ603775 MI-P-CP1-
    28      990        40.3    664  10  BE986058                BE986058 UI-M-CG0p
    29      980        39.9    621  10  BE986074                BE986074 UI-M-CG0p
    30      973        39.6    920  10  BF308306                BF308306 601890370
    31      949        38.6   2125  11  AK089246                AK089246 Mus muscu
    32      948        38.6    677  12  BG707107                BG707107 602670234
    33    947.5        38.6    684  12  BI907581                BI907581 603065948
    34      938        38.2    735  12  BM008943                BM008943 603618748
    35    933.5        38.0   2081  11  AK079650                AK079650 Mus muscu
    36      931        37.9    994  13  BQ644799                BQ644799 AGENCOURT
    37    912.5        37.1    683  13  BY750238                BY750238 BY750238
    38      900        36.6    638  10  BB612439                BB612439 BB612439
    39      889        36.2    608  10  BG079217                BG079217 H3037C03-
    40    888.5        36.2    908  14  CA327128                CA327128 UI-M-FY0-
    41      870        35.4    974  13  BQ944156                BQ944156 AGENCOURT
    42      869        35.4    594  12  BI402088                BI402088 MI-P-CP0-
    43      839        34.1    917  10  BF140667                BF140667 601786917
    44    837.5        34.1    584  14  CB586069                CB586069 AMGNNUC:N
    45    834.5        34.0    647  14  CA322312                CA322312 UI-M-FX0-


                                 ALIGNMENTS

RESULT 1
AK081806
LOCUS       AK081806              2101 bp    mRNA    linear   HTC 20-SEP-2003
DEFINITION  Mus musculus 16 days embryo head cDNA, RIKEN full-length enriched
            library, clone:C130078K14 product:solute carrier family 2,
            (facilitated glucose transporter), member 8, full insert sequence.
ACCESSION   AK081806
VERSION     AK081806.1  GI:26349396
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
```

```
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5
  AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6  (bases 1 to 2101)
  AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
            Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
            Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
            Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
            Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
            Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Please visit our web site for further details.
            URL:http://genome.gsc.riken.go.jp/
            URL:http://fantom.gsc.riken.go.jp/.
FEATURES             Location/Qualifiers
     source          1..2101
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:C130078K14"
                     /db_xref="MGI:2414880"
                     /db_xref="taxon:10090"
                     /clone="C130078K14"
                     /tissue_type="head"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="16 days embryo"
     CDS             61..1494
                     /note="unnamed protein product; putative
                     solute carrier family 2, (facilitated glucose
                     transporter), member 8 (MGD|MGI:1860103, GB|NM_019488,
                     evidence: BLASTN, 99%, match=1490)"
                     /codon_start=1
                     /protein_id="BAC38338.1"
                     /db_xref="GI:26349397"
                     /translation="MSPEDPQETQPLLRPPEARTPRGRRVFLASFAAALGPLSFGFAL
                     GYSSPAIPSLRRTAPPALRLGDNAASWFGAVVTLGAAAGGILGGWLLDRAGRKLSLLL
                     CTVPFVTGFAVITAARDVWMLLGGRLLTGLACGVASLVAPVYISEIAYPAVRGLLGSC
                     VQLMVVTGILLAYVAGWVLEWRWLAVLGCVPPTLMLLLMCYMPETPRFLLTQHQYQEA
                     MAALRFLWGSEEGWEEPPVGAEHQGFQLALLRRPGIYKPLIIGISLMVFQQLSGVNAI
                     MFYANSIFEEAKFKDSSLASVTVGIIQVLFTAVAALIMDRAGRRLLLALSGVIMVFSM
                     SAFGTYFKLTQSLPSNSSHVGLVPIAAEPVDVQVGLAWLAVGSMCLFIAGFAVGWGPI
                     PWLLMSEIFPLHVKGVATGICVLTNWFMAFLVTKEFSSVMEMLRPYGAFWLTAAFCAL
                     SVLFTLTVVPETKGRTLEQVTAHFEGR"
     polyA_signal    2080..2085
                     /note="putative"
     polyA_site      2101
                     /note="putative"
ORIGIN

Alignment Scores:
Pred. No.:              3.49e-185      Length:        2101
Score:                  2159.00        Matches:       412
Percent Similarity:     92.47%         Conservative:  30
Best Local Similarity:  86.19%         Mismatches:    34
Query Match:            87.87%         Indels:        2
DB:                     11             Gaps:          2

US-09-886-954A-1 (1-477) x AK081806 (1-2101)

Qy          1 MetThrProGluAspProGluGluThrGlnProLeuLeuGlyProProGlyGlySerAla 20
              |||:::||||||||||||:::||||||||||||||||||   ||||||
Db         61 ATGTCTCCCGAGGACCCCCAGGAGACGCAGCCGCTATTGCGGCCACCGGAAGCCAGGACT 120

Qy         21 ProArgGlyArgArgValPheLeuAlaAlaPheAlaAlaAlaLeuGlyProLeuSerPhe 40
              |||||||||||||||||||||||||||:::||||||||||||||||||||||||||||||
Db        121 CCCCGCGGCCGCCGGGTCTTCCTCGCTTCCTTCGCCGCCGCTCTGGGACCCCTCAGCTTC 180

Qy         41 GlyPheAlaLeuGlyTyrSerSerProAlaIleProSerLeuGlnArgAlaAlaProPro 60
              ||||||||||||||||||||||||||||||||||||||||||:::|||   |||||||||
Db        181 GGCTTCGCGCTCGGCTACAGCTCCCCCGCCATCCCCAGCCTGCGGCGCACCGCACCCCCG 240

Qy         61 AlaProArgLeuAspAspAlaAlaAlaSerTrpPheGlyAlaValValThrLeuGlyAla 80
              |||   ||||||   |||   |||||||||||||||||||||||||||||||||||||||
Db        241 GCCCTGCGCCTCGGAGACAATGCGGCCTCCTGGTTCGGGGCCGTCGTGACCCTGGGCGCT 300

Qy         81 AlaAlaGlyGlyValLeuGlyGlyTrpLeuValAspArgAlaGlyArgLysLeuSerLeu 100
              ||||||||||||:::|||||||||||||||:::|||||||||||||||||||||||||||
Db        301 GCTGCAGGGGGCATACTGGGCGGCTGGCTCCTGGACCGTGCAGGGCGCAAGCTGAGCCTC 360

Qy        101 LeuLeuCysSerValProPheValAlaGlyPheAlaValIleThrAlaAlaGlnAspVal 120
              |||||||||:::||||||||||||   ||||||||||||||||||||||||:::||||||
Db        361 TTGCTCTGCACCGTGCCCTTCGTGACTGGCTTTGCTGTCATCACCGCGGCCCGGGATGTG 420

Qy        121 TrpMetLeuLeuGlyGlyArgLeuLeuThrGlyLeuAlaCysGlyValAlaSerLeuVal 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TGGATGCTGCTCGGAGGCCGCCTCCTCACCGGCCTTGCCTGCGGAGTCGCCTCACTAGTG 480

Qy        141 AlaProValTyrIleSerGluIleAlaTyrProAlaValArgGlyLeuLeuGlySerCys 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GCACCGGTTTACATCTCGGAAATCGCCTACCCAGCTGTCCGAGGACTGCTCGGCTCCTGT 540

Qy        161 ValGlnLeuMetValValValGlyIleLeuLeuAlaTyrLeuAlaGlyTrpValLeuGlu 180
              ||||||||||||||||||   ||||||||||||||||||:::||||||||||||||||||
```

```
Db          541 GTGCAGCTGATGGTTGTCACTGGCATCCTCCTGGCCTATGTGGCAGGCTGGGTCCTAGAG 600

Qy          181 TrpArgTrpLeuAlaValLeuGlyCysValProProSerLeuMetLeuLeuLeuMetCys 200
                ||||||||||||||||||||||||||||||||||||:::|||||||||||||||||||||
Db          601 TGGCGCTGGCTGGCCGTGCTGGGCTGTGTGCCCCCCACCCTCATGCTGCTGCTCATGTGC 660

Qy          201 PheMetProGluThrProArgPheLeuLeuThrGlnHisArgArgGlnGluAlaMetAla 220
                :::||||||||||||||||||||||||||||||||||||:::   |||||||||||||||
Db          661 TACATGCCCGAGACCCCACGTTTTCTCCTCACTCAACACCAGTACCAGGAGGCCATGGCT 720

Qy          221 AlaLeuArgPheLeuTrpGlySerGluGlnGlyTrpGluAspProProIleGlyAlaGlu 240
                |||||||||||||||||||||||||||:::|||||||||:::||||||:::|||||||||
Db          721 GCCTTGCGCTTCCTGTGGGGCTCTGAGGAGGGCTGGGAAGAGCCCCCTGTTGGGGCTGAG 780

Qy          241 ---GlnSerPheHisLeuAlaLeuLeuArgGlnProGlyIleTyrLysProPheIleIle 259
                   |||   |||   |||||||||||||||:::||||||||||||||||||   ||||||
Db          781 CACCAGGGCTTCCAGCTGGCCCTGCTGAGGCGCCCTGGCATCTACAAGCCCCTCATCATC 840

Qy          260 GlyValSerLeuMetAlaPheGlnGlnLeuSerGlyValAsnAlaValMetPheTyrAla 279
                |||:::|||||||||   |||||||||||||||||||||||||||:::||||||||||||
Db          841 GGCATTTCCCTCATGGTCTTCCAGCAGCTGTCAGGGGTCAATGCTATCATGTTCTATGCC 900

Qy          280 GluThrIlePheGluGluAlaLysPheLysAspSerSerLeuAlaSerValValValGly 299
                   :::|||||||||||||||||||||||||||||||||||||||||||||   ||||||
Db          901 AACAGCATCTTCGAGGAGGCCAAGTTCAAGGACAGCAGCCTGGCCTCGGTCACTGTGGGC 960

Qy          300 ValIleGlnValLeuPheThrAlaValAlaAlaLeuIleMetAspArgAlaGlyArgArg 319
                :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          961 ATAATCCAGGTCCTGTTCACTGCTGTGGCGGCCCTCATCATGGACAGAGCAGGGCGAAGG 1020

Qy          320 LeuLeuLeuValLeuSerGlyValValMetValPheSerThrSerAlaPheGlyAlaTyr 339
                |||||||||   ||||||||||||:::||||||||||||   ||||||||||||   |||
Db         1021 CTGCTCCTGGCCTTGTCGGGTGTGATCATGGTGTTCAGTATGAGTGCCTTTGGTACCTAC 1080

Qy          340 PheLysLeuThrGlnGlyGlyProGlyAsnSerSerHisValAlaIleSerAlaProVal 359
                |||||||||||||||      |||   |||||||||||||||   :::      |||:::
Db         1081 TTCAAACTGACCCAGAGCCTCCCCAGCAACTCCTCCCACGTAGGCCTG---GTGCCCATC 1137

Qy          360 SerAlaGlnProValAspAlaSerValGlyLeuAlaTrpLeuAlaValGlySerMetCys 379
                :::|||:::|||||||||      ||||||||||||||||||||||||||||||||||||
Db         1138 GCGGCGGAGCCTGTGGATGTCCAAGTGGGACTGGCCTGGCTGGCTGTAGGCAGCATGTGC 1197

Qy          380 LeuPheIleAlaGlyPheAlaValGlyTrpGlyProIleProTrpLeuLeuMetSerGlu 399
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1198 CTCTTCATTGCTGGCTTTGCGGTGGGCTGGGGACCCATCCCCTGGCTCCTCATGTCAGAG 1257

Qy          400 IlePheProLeuHisValLysGlyValAlaThrGlyIleCysValLeuThrAsnTrpLeu 419
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1258 ATCTTCCCTCTGCATGTCAAGGGTGTGGCTACCGGCATCTGTGTCCTCACCAACTGGTTC 1317

Qy          420 MetAlaPheLeuValThrLysGluPheSerSerLeuMetGluValLeuArgProTyrGly 439
                |||||||||||||||||||||||||||||||||:::||||||:::|||||||||||||||
Db         1318 ATGGCCTTTCTAGTGACCAAAGAGTTCAGCAGCGTCATGGAGATGCTCAGACCCTACGGT 1377

Qy          440 AlaPheTrpLeuAlaSerAlaPheCysIlePheSerValLeuPheThrLeuPheCysVal 459
                ||||||||||||   :::|||||||||      ||||||||||||||||||      |||
Db         1378 GCCTTCTGGCTCACCGCTGCCTTCTGCGCTCTCAGTGTCCTATTCACACTGACCGTTGTC 1437

Qy          460 ProGluThrLysGlyLysThrLeuGluGlnIleThrAlaHisPheGluGlyArg 477
                |||||||||||||||:::||||||||||||:::|||||||||||||||||||||
Db         1438 CCTGAGACTAAAGGCAGGACTCTGGAACAAGTCACAGCCCATTTCGAGGGACGA 1491

RESULT 2
AF289587
LOCUS       AF289587                1954 bp    mRNA    linear   HTC 01-JAN-2002
DEFINITION  Homo sapiens clone pp7425 unknown mRNA.
ACCESSION   AF289587
VERSION     AF289587.1  GI:18027377
KEYWORDS    HTC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
REFERENCE   1  (bases 1 to 1954)
  AUTHORS   Zhang,P.P., Zhou,X.M., Jiang,H.Q., Huang,Y., Qin,W.X., Zhao,X.T.,
            Wan,D.F. and Gu,J.R.
  TITLE     Novel human cDNA clones with function of inhibiting cancer cell
            growth
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1954)
  AUTHORS   Zhang,P.P., Zhou,X.M., Jiang,H.Q., Huang,Y., Qin,W.X., Zhao,X.T.,
            Wan,D.F. and Gu,J.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-JUL-2000) National Laboratory For Oncogenes & Related
            Genes, Shanghai Cancer Institute, 25/Ln 2200 Xie-Tu Road, Shanghai
            200032, P. R. China
FEATURES             Location/Qualifiers
     source          1. .1954
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="pp7425"
     CDS             491. .1237
                     /codon_start=1
                     /product="unknown"
                     /protein_id="AAL55771.1"
                     /db_xref="GI:18027378"
                     /translation="MVVVGILLAYLAGWVLEWRWLAVLGCVPPSLMLLLMCFMPETPR
                     FLLTQHRRQEAMAALRFLWGSEQGWEDPPIGAEQSFHLALLRQPGIYKPFIIGVSLMA
                     FQQLSGVNAVMFYAETIFEEAKFKDSSLASVVVGVIQVLFTAVAALIMDRAGRRLLLV
                     LSGVVMVFSTSAFGAYFKLTQGGPGNSSHVAISAPVSAQPVDASVGLAWLAVGSMCLF
                     IAGGPQALWSLLACLRFLHLQCPFHFVLCP"
ORIGIN

Alignment Scores:
Pred. No.:              5.29e-182       Length:         1954
Score:                  2123.50         Matches:        426
Percent Similarity:     89.14%          Conservative:   1
Best Local Similarity:  88.94%          Mismatches:     1
Query Match:            86.43%          Indels:         51
DB:                     11              Gaps:           1

US-09-886-954A-1 (1-477) x AF289587 (1-1954)

Qy            1 MetThrProGluAspProGluGluThrGlnProLeuLeuGlyProProGlyGlySerAla 20
                :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db            1 GTGACGCCCGAGGACCCAGAGGAAACCCAGCCGCTTCTGGGGCCTCCTGGCGGCAGCGCG 60

Qy           21 ProArgGlyArgArgValPheLeuAlaAlaPheAlaAlaAlaLeuGlyProLeuSerPhe 40
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           61 CCCCGCGGCCGCCGCGTCTTCCTCGCCGCCTTCGCCGCTGCCCTGGGCCCACTCAGCTTC 120

Qy           41 GlyPheAlaLeuGlyTyrSerSerProAlaIleProSerLeuGlnArgAlaAlaProPro 60
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          121 GGCTTCGCGCTCGGCTACAGCTCCCCGGCCATCCCTAGCCTGCAGCGCGCCGCGCCCCCG 180

Qy           61 AlaProArgLeuAspAspAlaAlaAlaSerTrpPheGlyAlaValValThrLeuGlyAla 80
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          181 GCCCCGCGCCTGGACGACGCCGCCGCCTCCTGGTTCGGGGCTGTCGTGACCCTGGGTGCC 240

Qy           81 AlaAlaGlyGlyValLeuGlyGlyTrpLeuValAspArgAlaGlyArgLysLeuSerLeu 100
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          241 GCGGCGGGGGGAGTGCTGGGCGGCTGGCTGGTGGACCGCGCCGGGCGCAAGCTGAGCCTC 300

Qy          101 LeuLeuCysSerValProPheValAlaGlyPheAlaValIleThrAlaAlaGlnAspVal 120
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          301 TTGCTGTGCTCCGTGCCCTTCGTGGCCGGCTTTGCCGTCATCACCGCGGCCCAGGACGTG 360

Qy          121 TrpMetLeuLeuGlyGlyArgLeuLeuThrGlyLeuAlaCysGlyValAlaSerLeuVal 140
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          361 TGGATGCTGCTGGGGGGCCGCCTCCTCACCGGCCTGGCCTGCGGTGTTGCCTCCCTAGTG 420

Qy          141 AlaProValTyrIleSerGluIleAlaTyrProAlaVal-ArgGlyLeuLeuGlySerCy 160
                ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db          421 GCCCCGGTCTACATCTCCGAAATCGCCTACCCAGCAGTCCCGGGGGTTGCTCGGCTCCTG 480
```

```
Qy        160 sValGlnLeuMetValValValGlyIleLeuLeuAlaTyrLeuAlaGlyTrpValLeuGl 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TGTGCAGCTAATGGTCGTCGTCGGCATCCTCCTGGCCTACCTGGCAGGCTGGGTGCTGGA 540

Qy        180 uTrpArgTrpLeuAlaValLeuGlyCysValProProSerLeuMetLeuLeuLeuMetCy 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTGGCGCTGGCTGGCTGTGCTGGGCTGCGTGCCCCCCTCCCTCATGCTGCTTCTCATGTG 600

Qy        200 sPheMetProGluThrProArgPheLeuLeuThrGlnHisArgArgGlnGluAlaMetAl 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTTCATGCCCGAGACCCCGCGCTTCCTGCTGACTCAGCACAGGCGCCAGGAGGCCATGGC 660

Qy        220 aAlaLeuArgPheLeuTrpGlySerGluGlnGlyTrpGluAspProProIleGlyAlaGl 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CGCCCTGCGGTTCCTGTGGGGCTCCGAGCAGGGCTGGGAAGACCCCCCCATCGGGGCTGA 720

Qy        240 uGlnSerPheHisLeuAlaLeuLeuArgGlnProGlyIleTyrLysProPheIleIleGl 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GCAGAGCTTTCACCTGGCCCTGCTGCGGCAGCCCGGCATCTACAAGCCCTTCATCATCGG 780

Qy        260 yValSerLeuMetAlaPheGlnGlnLeuSerGlyValAsnAlaValMetPheTyrAlaGl 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CGTCTCCCTGATGGCCTTCCAGCAGCTGTCGGGGGTCAACGCCGTCATGTTCTATGCAGA 840

Qy        280 uThrIlePheGluGluAlaLysPheLysAspSerSerLeuAlaSerValValValGlyVa 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GACCATCTTTGAAGAGGCCAAGTTCAAGGACAGCAGCCTGGCCTCGGTCGTCGTGGGTGT 900

Qy        300 lIleGlnValLeuPheThrAlaValAlaAlaLeuIleMetAspArgAlaGlyArgArgLe 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 CATCCAGGTGCTGTTCACAGCTGTGGCGGCTCTCATCATGGACAGAGCAGGGCGGAGGCT 960

Qy        320 uLeuLeuValLeuSerGlyValValMetValPheSerThrSerAlaPheGlyAlaTyrPh 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GCTCCTGGTCTTGTCAGGTGTGGTCATGGTGTTCAGCACGAGTGCCTTCGGCGCCTACTT 1020

Qy        340 eLysLeuThrGlnGlyGlyProGlyAsnSerSerHisValAlaIleSerAlaProValSe 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CAAGCTGACCCAGGGTGGCCCTGGCAACTCCTCGCACGTGGCCATCTCGGCGCCTGTCTC 1080

Qy        360 rAlaGlnProValAspAlaSerValGlyLeuAlaTrpLeuAlaValGlySerMetCysLe 380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 TGCACAGCCTGTTGATGCCAGCGTGGGGCTGGCCTGGCTGGCCGTGGGCAGCATGTGCCT 1140

Qy        380 uPheIleAlaGly-PheAlaValGlyTrpGlyProIleProTrpLeuLeuMetSerGluI 400
              |||||||||||||
Db       1141 CTTCATCGCCGGAGGTC------------------------------------------- 1157

Qy        400 lePheProLeuHisValLysGlyValAlaThrGlyIleCysValLeuThrAsnTrpLeuM 420

Db       1157 ------------------------------------------------------------ 1157

Qy        420 etAlaPheLeuValThrLysGluPheSerSerLeuMetGluValLeuArgProTyrGlyA 440
                                                          ||||||||||||||||
Db       1158 --------------------------------------------CTCAGGCCCTATGGAG 1173

Qy        440 laPheTrpLeuAlaSerAlaPheCysIlePheSerValLeuPheThrLeuPheCysValP 460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1174 CCTTCTGGCTTGCCTCCGCTTTCTGCATCTTCAGTGTCCTTTTCACTTTGTTCTGTGTCC 1233

Qy        460 roGluThrLysGlyLysThrLeuGluGlnIleThrAlaHisPheGluGlyArg 477
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1234 CTGAAACTAAAGGAAAGACTCTGGAACAAATCACAGCCCATTTTGAGGGGCGA 1286

RESULT 3
AY414180
LOCUS       AY414180                1008 bp    DNA     linear   GSS 17-DEC-2003
DEFINITION  Homo sapiens SLC2A8 gene, VIRTUAL TRANSCRIPT, partial sequence,
            genomic survey sequence.
ACCESSION   AY414180
VERSION     AY414180.1  GI:39770142
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1008)
  AUTHORS   Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
            Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
            Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
            Adams,M.D. and Cargill,M.
  TITLE     Inferring nonneutral evolution from human-chimp-mouse orthologous
            gene trios
  JOURNAL   Science 302 (5652), 1960-1963 (2003)
   PUBMED   14671302
REFERENCE   2  (bases 1 to 1008)
  AUTHORS   Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
            Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
            Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
            Adams,M.D. and Cargill,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
            Rockville, MD 20850, USA
COMMENT     This sequence was made by sequencing genomic exons and ordering
            them based on alignment.
FEATURES             Location/Qualifiers
     source          1..1008
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
     gene            <1..>1008
                     /gene="SLC2A8"
                     /locus_tag="HCM5138"
ORIGIN

Alignment Scores:
Pred. No.:              4.13e-145     Length:        1008
Score:                  1713.00       Matches:       332
Percent Similarity:     99.10%        Conservative:  0
Best Local Similarity:  99.10%        Mismatches:    3
Query Match:            69.72%        Indels:        0
DB:                     29            Gaps:          0

US-09-886-954A-1 (1-477) x AY414180 (1-1008)

Qy        143 ValTyrIleSerGluIleAlaTyrProAlaValArgGlyLeuLeuGlySerCysValGln 162
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTCTACATCTCCGAAATCGCCTACCCAGCAGTCCGGGGGTTGCTCGGCTCCTGTGTGCAG 60

Qy        163 LeuMetValValValGlyIleLeuLeuAlaTyrLeuAlaGlyTrpValLeuGluTrpArg 182
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CTAATGGTCGTCGTCGGCATCCTCCTGGCCTACCTGGCAGGCTGGGTGCTGGAGTGGCGC 120

Qy        183 TrpLeuAlaValLeuGlyCysValProProSerLeuMetLeuLeuLeuMetCysPheMet 202
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TGGCTGGCTGTGCTGGGCTGCGTGCCCCCCTCCCTCATGCTGCTTCTCATGTGCTTCATG 180

Qy        203 ProGluThrProArgPheLeuLeuThrGlnHisArgArgGlnGluAlaMetAlaAlaLeu 222
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCCGAGACCCCGCGCTTCCTGCTGACTCAGCACAGGCGCCAGGAGGCCATGGCCGCCCTG 240

Qy        223 ArgPheLeuTrpGlySerGluGlnGlyTrpGluAspProProIleGlyAlaGluGlnSer 242
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CGGTTCCTGTGGGGCTCCGAGCAGGGCTGGGAAGACCCCCCCATCGGGGCTGAGCAGAGC 300

Qy        243 PheHisLeuAlaLeuLeuArgGlnProGlyIleTyrLysProPheIleIleGlyValSer 262
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TTTCACCTGGCCCTGCTGCGGCAGCCCGGCATCTACAAGCCCTTCATCATCGGCGTCTCC 360

Qy        263 LeuMetAlaPheGlnGlnLeuSerGlyValAsnAlaValMetPheTyrAlaGluThrIle 282
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CTGATGGCCTTCCAGCAGCTGTCGGGGGTCAACGCCGTCATGTTCTATGCAGAGACCATC 420

Qy        283 PheGluGluAlaLysPheLysAspSerSerLeuAlaSerValValValGlyValIleGln 302
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TTTGAAGAGGCCAAGTTCAAGGACAGCAGCCTGGCCTCGGTCGTCGTGGGTGTCATCCAG 480
```

```
Qy        303 ValLeuPheThrAlaValAlaAlaLeuIleMetAspArgAlaGlyArgArgLeuLeuLeu 322
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GTGCTGTTCACAGCTGTGGCGGCTCTCATCATGGACAGAGCAGGGCGGAGGCTGCTCCTG 540

Qy        323 ValLeuSerGlyValValMetValPheSerThrSerAlaPheGlyAlaTyrPheLysLeu 342
              |||         ||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTCTTNNNNNGTGTGGTCATGGTGTTCAGCACGAGTGCCTTCGGCGCCTACTTCAAGCTG 600

Qy        343 ThrGlnGlyGlyProGlyAsnSerSerHisValAlaIleSerAlaProValSerAlaGln 362
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 ACCCAGGGTGGCCCTGGCAACTCCTCGCACGTGGCCATCTCGGCGCCTGTCTCTGCACAG 660

Qy        363 ProValAspAlaSerValGlyLeuAlaTrpLeuAlaValGlySerMetCysLeuPheIle 382
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CCTGTTGATGCCAGCGTGGGGCTGGCCTGGCTGGCCGTGGGCAGCATGTGCCTCTTCATC 720

Qy        383 AlaGlyPheAlaValGlyTrpGlyProIleProTrpLeuLeuMetSerGluIlePhePro 402
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GCCGGCTTTGCGGTGGGCTGGGGGCCCATCCCCTGGCTCCTCATGTCAGAGATCTTCCCT 780

Qy        403 LeuHisValLysGlyValAlaThrGlyIleCysValLeuThrAsnTrpLeuMetAlaPhe 422
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CTGCATGTCAAGGGCGTGGCGACAGGCATCTGCGTCCTCACCAACTGGCTCATGGCCTTT 840

Qy        423 LeuValThrLysGluPheSerSerLeuMetGluValLeuArgProTyrGlyAlaPheTrp 442
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CTCGTGACCAAGGAGTTCAGCAGCCTCATGGAGGTCCTCAGGCCCTATGGAGCCTTCTGG 900

Qy        443 LeuAlaSerAlaPheCysIlePheSerValLeuPheThrLeuPheCysValProGluThr 462
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 CTTGCCTCCGCTTTCTGCATCTTCAGTGTCCTTTTCACTTTGTTCTGTGTCCCTGAAACT 960

Qy        463 LysGlyLysThrLeuGluGlnIleThrAlaHisPheGluGlyArg 477
              |||||||||||||||||||||||||||||||||||||||||||||
Db        961 AAAGGAAAGACTCTGGAACAAATCACAGCCCATTTTGAGGGGCGA 1005

RESULT 4
BX395379
LOCUS       BX395379                1201 bp    mRNA    linear   EST 13-MAY-2003
DEFINITION  BX395379 Homo sapiens NEUROBLASTOMA COT 50-NORMALIZED Homo sapiens
            cDNA clone CS0DD006YP06 5-PRIME, mRNA sequence.
ACCESSION   BX395379
VERSION     BX395379.1  GI:30620615
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1201)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 5383.f For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DD006DH03QP1&cluster=5383.f. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DD006DH03QP1.
FEATURES             Location/Qualifiers
     source          1..1201
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DD006YP06"
                     /tissue_type="NEUROBLASTOMA COT 50-NORMALIZED"
                     /clone_lib="Homo sapiens NEUROBLASTOMA COT 50-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN

Alignment Scores:
Pred. No.:              9.65e-125      Length:         1201
Score:                  1490.50        Matches:        320
Percent Similarity:     68.72%         Conservative:   3
Best Local Similarity:  68.09%         Mismatches:     23
Query Match:            60.66%         Indels:         127
DB:                     13             Gaps:           1

US-09-886-954A-1 (1-477) x BX395379 (1-1201)

Qy          1 MetThrProGluAspProGluGluThrGlnProLeuLeuGlyProProGlyGlySerAla 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         83 ATGACGCCCGAGGACCCAGAGGAAACCCAGCCGCTTCTGGGGCCTCCTGGCGGCAG---- 138

Qy         21 ProArgGlyArgArgValPheLeuAlaAlaPheAlaAlaAlaLeuGlyProLeuSerPhe 40

Db        138 ------------------------------------------------------------ 138

Qy         41 GlyPheAlaLeuGlyTyrSerSerProAlaIleProSerLeuGlnArgAlaAlaProPro 60

Db        138 ------------------------------------------------------------ 138

Qy         61 AlaProArgLeuAspAspAlaAlaAlaSerTrpPheGlyAlaValValThrLeuGlyAla 80

Db        138 ------------------------------------------------------------ 138

Qy         81 AlaAlaGlyGlyValLeuGlyGlyTrpLeuValAspArgAlaGlyArgLysLeuSerLeu 100

Db        138 ------------------------------------------------------------ 138

Qy        101 LeuLeuCysSerValProPheValAlaGlyPheAlaValIleThrAlaAlaGlnAspVal 120

Db        138 ------------------------------------------------------------ 138

Qy        121 TrpMetLeuLeuGlyGlyArgLeuLeuThrGlyLeuAlaCysGlyValAlaSerLeuVal 140

Db        138 ------------------------------------------------------------ 138

Qy        141 AlaProValTyrIleSerGluIleAlaTyrProAlaValArgGlyLeuLeuGlySerCys 160
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        139 ------GTCTACATCTCCGAAATCGCCTACCCAGCAGTWCGGGGGTTGCTCGGCTCCTGT 192

Qy        161 ValGlnLeuMetValValValGlyIleLeuLeuAlaTyrLeuAlaGlyTrpValLeuGlu 180
              ||||||||||||||||||||||||||||||   ||||||||||||||||||     |||
Db        193 GTGCAGCTAATGGTCGTCGTCGGCATCCTATG-GCCTACCTGGCAGGCTGGKTGHTGGAG 251

Qy        181 TrpArgTrpLeuAlaValLeuGlyCysValProProSerLeuMetLeuLeuLeuMetCys 200
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        252 TKGCGCTGGCTGGCTGTGCTGGGCTGCGTGCCCCCCTCCCTCATGCTGCTTCTCATGTGC 311

Qy        201 PheMetProGluThrProArgPheLeuLeuThrGlnHisArgArgGlnGluAlaMetAla 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        312 TTTATGCCCGAGACCCCGCGCTTCCTGCTGACTCAGCACAGGCGCCAGGAGGCCATGGCC 371

Qy        221 AlaLeuArgPheLeuTrpGlySerGluGlnGlyTrpGluAspProProIleGlyAlaGlu 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        372 GCCCTGCGNTTCCTGTGGGGCTCCGAGCAGGGCTGGGAAGACCCCCCCATCGGGGCTGAG 431

Qy        241 GlnSerPheHisLeuAlaLeuLeuArgGlnProGlyIleTyrLysProPheIleIleGly 260
              ||||||||||||||||||||||||||||||   |||||||||||||      ||||||||
Db        432 CAGAGCTTTCACCTGGCCCTGCTGCGGCAGCNCGGCATCTACAAGCYCTTHATCATCGGC 491

Qy        261 ValSerLeuMetAlaPheGlnGlnLeuSerGlyValAsnAlaValMetPheTyrAlaGlu 280
              |||||||||||||||||||||||||||   ||||||||||||||||||||||||||||||
Db        492 GTCTCCCTGATGGCCTTCCAGCAGCTGTTGGGGGTTAACGCCGTTATGTTCTATGCAGAG 551

Qy        281 ThrIlePheGluGluAlaLysPheLysAspSerSerLeuAlaSerValValValGlyVal 300
```

```
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         552 ACCATCTTTGAAGAGGCCAAGTTCAAGGACAGCAGCCTGGCCTCGGTCGTCGTGGGTGTY 611

Qy         301 IleGlnValLeuPheThrAlaValAlaAlaLeuIleMetAspArgAlaGlyArgArgLeu 320
               ||||||||||||   |||||||||||||||||||||||||||||||||||||||||||||
Db         612 ATCCAGGTGCTGTWCACAGCTGTAGCGGCTCTYATCATGGACAGAGCAGGGCGGAGGCTG 671

Qy         321 LeuLeuValLeuSerGlyValValMetValPheSerThrSerAlaPheGlyAlaTyrPhe 340
               ||||||||||||   |||||||||||||||||||||||||||:::|||||||||||||||
Db         672 CTCCTGGTCTTGTYAGGTGTGGTCATGGTGTTCAGCACGAGTTCCTTCGGCGCCTACTTC 731

Qy         341 LysLeuThrGlnGlyGlyProGlyAsnSerSerHisValAlaIleSerAlaProValSer 360
               |||||||||||||||||||||||||||||||||||||||||||||   ||||||||||||
Db         732 AAGCTGACCCAGGGTGGCCCTGGCAACTCCTCGCACGTWGCCATMTMGGCGCCTGTTTCT 791

Qy         361 AlaGlnProValAspAlaSerValGlyLeuAlaTrpLeuAlaValGlySerMetCysLeu 380
               |||||||||||||||||||||   ||||||||||||||||||||||||||||||||||||
Db         792 GCACAGCCTGTTGATGCCAGCRTGGGGCTGGCCTGGCTGGCCGTGGGCAGCATGTGCCTW 851

Qy         381 PheIleAlaGlyPheAlaValGlyTrpGlyProIleProTrpLeuLeuMetSerGluIle 400
                  ||||||||||||||||||||||||||||||||||||||||||||||||   ||||||
Db         852 TWCATCGCCGGCTTTGCGGTGGGCTGGGGGCCCATCCCCTGGCTCCTCATGTHAGAAATC 911

Qy         401 PheProLeuHisValLysGlyValAlaThrGlyIleCysValLeuThrAsnTrpLeuMet 420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         912 TTCCCTCTGCATGTTAAGGGCGTKGCGACAGGVATCTGCGTTCTCACCAACTGGCTCATG 971

Qy         421 AlaPheLeuValThrLysGluPheSerSerLeuMetGluValLeuArgProTyrGlyAla 440
               ||||||||||||||||||||||||||||||   ||||||||||||||||||||||||||||||
Db         972 GCCTTTCTCGTGACCAAGGAGTTCAGCAGCTTCATGGAGGTCCTCAGGCCCTATGGAGCC 1031

Qy         441 PheTrpLeuAlaSerAlaPheCysIlePheSerValLeuPheThrLeuPheCysValPro 460
               |||   |||   |||||||||||||||   |||   |||||||||||||||      |||
Db        1032 TTCTKGCTTKCCTCCGCTTTCTGCATCTCCAGTTT-CTTTTCACTTTGTTCTKTTGTCCT 1090

Qy         461 -GluThrLysGlyLysThrLeuGluGln 469
                :::||||||||||||   |||:::|||
Db        1091 TAAAACTAAAGGAAARATTTTGGACCAA 1118

RESULT 5
AY414182
LOCUS       AY414182                1000 bp    DNA     linear   GSS 17-DEC-2003
DEFINITION  Mus musculus SLC2A8 gene, VIRTUAL TRANSCRIPT, partial sequence,
            genomic survey sequence.
ACCESSION   AY414182
VERSION     AY414182.1  GI:39770144
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1000)
  AUTHORS   Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
            Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
            Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
            Adams,M.D. and Cargill,M.
  TITLE     Inferring nonneutral evolution from human-chimp-mouse orthologous
            gene trios
  JOURNAL   Science 302 (5652), 1960-1963 (2003)
   PUBMED   14671302
REFERENCE   2  (bases 1 to 1000)
  AUTHORS   Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
            Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
            Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
            Adams,M.D. and Cargill,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
            Rockville, MD 20850, USA
COMMENT     This sequence was made by sequencing genomic exons and ordering
            them based on alignment.
FEATURES             Location/Qualifiers
     source          1..1000
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
     gene            <1..>1000
                     /gene="SLC2A8"
                     /locus_tag="HCM5138"
ORIGIN

Alignment Scores:
Pred. No.:              2.05e-124     Length:         1000
Score:                  1485.50       Matches:        284
Percent Similarity:     91.07%        Conservative:   22
Best Local Similarity:  84.52%        Mismatches:     26
Query Match:            60.46%        Indels:         4
DB:                     29            Gaps:           2

US-09-886-954A-1 (1-477) x AY414182 (1-1000)

Qy         143 ValTyrIleSerGluIleAlaTyrProAlaValArgGlyLeuLeuGlySerCysValGln 162
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 GTTTACATCTCGGAAATCGCCTACCCAGCTGTCCGAGGACTGCTCGGCTCCTGTGTGCAG 60

Qy         163 LeuMetValValValGlyIleLeuLeuAlaTyrLeuAlaGlyTrpValLeuGluTrpArg 182
               ||||||||||||   ||||||||||||||||||:::||||||||||||||||||||||||
Db          61 CTGATGGTTGTCACTGGCATCCTCCTGGCCTATGTGGCAGGCTGGGTCCTAGAGTGGCGC 120

Qy         183 TrpLeuAlaValLeuGlyCysValProProSerLeuMetLeuLeuLeuMetCysPheMet 202
               ||||||||||||||||||||||||||||||:::|||||||||||||||||||||:::|||
Db         121 TGGCTGGCCGTGCTGGGCTGTGTGCCCCCCACCCTCATGCTGCTGCTCATGTGCTACATG 180

Qy         203 ProGluThrProArgPheLeuLeuThrGlnHisArgArgGlnGluAlaMetAlaAlaLeu 222
               |||||||||||||||||||||||||||||||||:::   |||||||||||||||||||||
Db         181 CCCGAGACCCCACGTTTTCTCCTCACTCAACACCAGTACCAGGAGGCCATGGCTGCCTTG 240

Qy         223 ArgPheLeuTrpGlySerGluGlnGlyTrpGluAspProProIleGlyAlaGluGlnSer 242
               |||||||||||||||||||||:::|||||||||:::||||||:::|||||||||      
Db         241 CGCTTCCTGTGGGGCTCTGAGGAGGGCTGGGAAGAGCCCCCTGTTGGGGCTGAGCACGGC 300

Qy         243 PheHisLeuAlaLeuLeuArgGlnProGlyIleTyrLysProPheIleIleGlyValSer 262
               |||   |||||||||||||||:::||||||||||||||||||   |||||||||:::|||
Db         301 TTCCAGCTGGCCCTGCTGAGGCGCCCTGGCATCTACAAGCCCCTCATCATCGGCATTTCC 360

Qy         263 LeuMetAlaPheGlnGlnLeuSerGlyValAsnAlaValMetPheTyrAlaGluThrIle 282
               ||||||   |||||||||||||||||||||||||||:::||||||||||||   :::|||
Db         361 CTCATGGTCTTCCAGCAGCTGTCAGGGGTCAATGCTATCATGTTCTATGCCAACAGCATC 420

Qy         283 PheGluGluAlaLysPheLysAspSerSer-LeuAlaSerValValValGlyValIleGl 302
               |||||||||||||||||||||          ||||||||||||   ||||||:::|||||
Db         421 TTCGAGGAGGCCAAGTTCAAG------CAGCCTGGCCTCGGTCACTGTGGGCATAATCCA 474

Qy         302 nValLeuPheThrAlaValAlaAlaLeuIleMetAspArgAlaGlyArgArgLeuLeuLe 322
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         475 GGTCCTGTTCACTGCTGTGGCGGCCCTCATCATGGACAGAGCAGGGCGAAGGCTGCTCCT 534

Qy         322 uValLeuSerGlyValValMetValPheSerThrSerAlaPheGlyAlaTyrPheLysLe 342
               |   ||||||||||||:::||||||||||||   ||||||||||||   |||||||||||
Db         535 GGCCTTGTCGGGTGTGATCATGGTGTTCAGTATGAGTGCCTTTGGTACCTACTTCAAACT 594

Qy         342 uThrGlnGlyGlyProGlyAsnSerSerHisValAlaIleSerAlaProValSerAlaGl 362
               |||||||      |||   |||||||||||||||   :::      |||::::::|||::
Db         595 GACCCAGAGCCTCCCCAGCAACTCCTCCCACGTAGGCCTG---GTGCCCATCGCGGCGGA 651

Qy         362 nProValAspAlaSerValGlyLeuAlaTrpLeuAlaValGlySerMetCysLeuPheIl 382
               :|||||||||      ||||||||||||||||||||||||||||||||||||||||||||
Db         652 GCCTGTGGATGTCCAAGTGGGACTGGCCTGGCTGGCTGTAGGCAGCATGTGCCTCTTCAT 711

Qy         382 eAlaGlyPheAlaValGlyTrpGlyProIleProTrpLeuLeuMetSerGluIlePhePr 402
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         712 TGCTGGCTTTGCGGTGGGCTGGGGACCCATCCCCTGGCTCCTCATGTCAGAGATCTTCCC 771

Qy         402 oLeuHisValLysGlyValAlaThrGlyIleCysValLeuThrAsnTrpLeuMetAlaPh 422
               |||||||||||||||||||||||||||||||||||||||||||||||||   ||||||||
Db         772 TCTGCATGTCAAGGGTGTGGCTACCGGCATCTGTGTCCTCACCAACTGGTTCATGGCCTT 831
```

```
Qy        422 eLeuValThrLysGluPheSerSerLeuMetGluValLeuArgProTyrGlyAlaPheTr 442
              |||||||||||||||||||||||||:::||||||:::|||||||||||||||||||||||
Db        832 TCTAGTGACCAAAGAGTTCAGCAGCGTCATGGAGATGCTCAGACCCTACGGTGCCTTCTG 891

Qy        442 pLeuAlaSerAlaPheCysIlePheSerValLeuPheThrLeuPheCysValProGluTh 462
              ||||   :::|||||||||      ||||||||||||||||||      |||||||||||
Db        892 GCTCACCGCTGCCTTCTGCGCTCTCAGTGTCCTATTCACACTGACCGTTGTCCCTGAGAC 951

Qy        462 rLysGlyLysThrLeuGluGlnIleThrAlaHisPheGluGlyArg 477
              |||||||:::||||||||||||:::|||||||||||||||||||||
Db        952 TAAAGGCAGGACTCTGGAACAAGTCACAGCCCATTTCGAGGGACGA 997

RESULT 6
BM545247
LOCUS       BM545247                1085 bp    mRNA    linear   EST 20-FEB-2002
DEFINITION  AGENCOURT_6497266 NIH_MGC_124 Homo sapiens cDNA clone IMAGE:5726945
            5', mRNA sequence.
ACCESSION   BM545247
VERSION     BM545247.1  GI:18777177
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1085)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Invitrogen
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12720  row: d  column: 18
            High quality sequence stop: 679.
FEATURES             Location/Qualifiers
     source          1..1085
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5726945"
                     /tissue_type="hippocampus"
                     /lab_host="DH10B"
                     /clone_lib="NIH_MGC_124"
                     /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: EcoRV
                     (destroyed); Site_2: NotI; RNA source male hippocampus,
                     age 27. Library is oligo-dT primed and directionally
                     cloned (EcoRV site is destroyed upon cloning). Average
                     insert size 1.4 kb, insert size range 0.9-4 kb. Library is
                     normalized and enriched for full-length clones and was
                     constructed by C. Gruber (Invitrogen). Research Genetics
                     tracking code 012."
ORIGIN

Alignment Scores:
Pred. No.:              1.15e-114       Length:         1085
Score:                  1379.00         Matches:        279
Percent Similarity:     87.77%          Conservative:   1
Best Local Similarity:  87.46%          Mismatches:     6
Query Match:            56.13%          Indels:         33
DB:                     12              Gaps:           3

US-09-886-954A-1 (1-477) x BM545247 (1-1085)

Qy        162 GlnLeuMetValValValGlyIleLeuLeuAlaTyrLeuAlaGlyTrpValLeuGluTrp 181
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CAGCTAATGGTCGTCGTCGGCATCCTCCTGGCCTACCTGGCAGGCTGGGTGCTGGAGTGG 60

Qy        182 ArgTrpLeuAlaValLeuGlyCysValProProSerLeuMetLeuLeuLeuMetCysPhe 201
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCTGGCTGGCTGTGCTGGGCTGCGTGCCCCCCTCCCTCATGCTGCTTCTCATGTGCTTC 120

Qy        202 MetProGluThrProArgPheLeuLeuThrGlnHisArgArgGlnGluAlaMetAlaAla 221
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ATGCCCGAGACCCCGCGCTTCCTGCTGACTCAGCACAGGCGCCAGGAGGCCATGGCCGCC 180

Qy        222 LeuArgPheLeuTrpGlySerGluGlnGlyTrpGluAspProProIleGlyAlaGluGln 241
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CTGCGGTTCCTGTGGGGCTCCGAGCAGGGCTGGGAAGACCCCCCCATCGGGGCTGAGCAG 240

Qy        242 SerPheHisLeuAlaLeuLeuArgGlnProGlyIleTyrLysProPheIleIleGlyVal 261
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGCTTTCACCTGGCCCTGCTGCGGCAGCCCGGCATCTACAAGCCCTTCATCATCGGCGTC 300

Qy        262 SerLeuMetAlaPheGlnGlnLeuSerGlyValAsnAlaValMetPheTyrAlaGluThr 281
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TCCCTGATGGCCTTCCAGCAGCTGTCGGGGGTCAACGCCGTCATGTTCTATGCAGAGACC 360

Qy        282 IlePheGluGluAlaLysPheLysAspSerSerLeuAlaSerValValValGlyValIle 301
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ATCTTTGAAGAGGCCAAGTTCAAGGACAGCAGCCTGGCCTCGGTCGTCGTGGGTGTCATC 420

Qy        302 GlnValLeuPheThrAlaValAlaAlaLeuIleMetAspArgAlaGlyArgArgLeuLeu 321
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CAGGTGCTGTTCACAGCTGTGGCGGCTCTCATCATGGACAGAGCAGGGCGGAGGCTGCTC 480

Qy        322 LeuValLeuSerGlyValValMetValPheSerThrSerAlaPheGlyAlaTyrPheLys 341
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CTGGTCTTGTCAGGTGTGGTCATGGTGTTCAGCACGAGTGCCTTCGGCGCCTACTTCAAG 540

Qy        342 LeuThrGlnGlyGlyProGlyAsnSerSerHisValAlaIleSerAlaProValSerAla 361
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CTGACCCAGGGTGGCCCTGGCAACTCCTCGCACGTGGCCATCTCGGCGCCTGTCTCTGCA 600

Qy        362 GlnProValAspAlaSerValGlyLeuAlaTrpLeuAlaValGlySerMetCysLeuPhe 381
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CAGCCTGTTGATGCCAGCGTGGGGCTGGCCTGGCTGGCCGTGGGCAGCATGTGCCTCTTC 660

Qy        382 IleAlaGlyPheAlaVal-GlyTrpGlyProIleProTrpLeuLeuMetSerGluIlePh 401
              |||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db        661 ATCGCCGGCTTTGCGGTTGGGCTGGGGGCCCATCCCCTGGCTCCTCATGTCAGAGATCTT 720

Qy        401 eProLeuHisValLysGlyValAlaThrGlyIleCysValLeuThrAsnTrpLeuMetAl 421
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCCTCTGCATGTCAAGGGCGTGGCGACAGGCATCTGCGTCCTCACCAACTGGCTCATGGC 780

Qy        421 aPheLeuValThrLysGluPheSerSerLeuMet-------------------------- 432
              |||||||   ||||||||||||||||||||||||
Db        781 CTTTCTCNTGACCAAGGAGTTCAGCAGCCTCATGCTCCTGCTGCTCTGAGGACTCAN 840

Qy        433 ----------GluValLeuArgPro----------------------------------- 437
                        |||:::   ||||||
Db        841 GAACACCTTCGAGCTTTGCAGACCTGCGGTCAGCCCTCCATGCGCAAGATAAAGCAACGG 900

Qy        438 -------------------TyrGly------AlaPheTrpLeuAlaSerAlaPhe 447
                                    |||      ||||||||||||         |||
Db        901 NAAAAAGGAGGGGGGCCTCCAAGGACCTTTGGCTTTCTGGCTTGAAGGGCCTTTT 955

RESULT 7
BG479842
LOCUS       BG479842                 932 bp    mRNA    linear   EST 21-MAR-2001
DEFINITION  602527373F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:4650906 5',
            mRNA sequence.
ACCESSION   BG479842
VERSION     BG479842.1  GI:13412121
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 932)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
```

```
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1433  row: a  column: 19
            High quality sequence stop: 820.
FEATURES             Location/Qualifiers
     source          1..932
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4650906"
                     /tissue_type="choriocarcinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_21"
                     /note="Organ: placenta; Vector: pOTB7; Site_1: XhoI;
                     Site_2: EcoRI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G).  Size-selected >500bp
                     for average insert size 1.8kb.  Library constructed by
                     Ling Hong in the laboratory of Gerald M. Rubin (University
                     of California, Berkeley) using  ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
ORIGIN

Alignment Scores:
Pred. No.:              3.18e-111      Length:          932
Score:                  1340.00        Matches:         293
Percent Similarity:     93.93%         Conservative:    1
Best Local Similarity:  93.61%         Mismatches:      17
Query Match:            54.54%         Indels:          7
DB:                     12             Gaps:            0

US-09-886-954A-1 (1-477) x BG479842 (1-932)

Qy         51 IleProSerLeuGlnArgAlaAlaProProAlaProArgLeuAspAspAlaAlaAlaSer 70
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 ATCCCTAGCCTGCAGCGCGCCGCGCCCCCGGCCCCGCGCCTGGACGACGCCGCCGCCTCC 61

Qy         71 TrpPheGlyAlaValValThrLeuGlyAlaAlaAlaGlyGlyValLeuGlyGlyTrpLeu 90
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 TGGTTCGGGGCTGTCGTGACCCTGGGTGCCGCGGCGGGGGGAGTGCTGGGCGGCTGGCTG 121

Qy         91 ValAspArgAlaGlyArgLysLeuSerLeuLeuLeuCysSerValProPheValAlaGly 110
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 GTGGACCGCGCCGGGCGCAAGCTGAGCCTCTTGCTGTGCTCCGTGCCCTTCGTGGCCGGC 181

Qy        111 PheAlaValIleThrAlaAlaGlnAspValTrpMetLeuLeuGlyGlyArgLeuLeuThr 130
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        182 TTTGCCGTCATCACCGCGGCCCAGGACGTGTGGATGCTGCTGGGGGGCCGCCTCCTCACC 241

Qy        131 GlyLeuAlaCysGlyValAlaSerLeuValAlaProValTyrIleSerGluIleAlaTyr 150
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        242 GGCCTGGCCTGCGGTGTTGCCTCCCTAGTGGCCCCGGTCTACATCTCCGAAATCGCCTAC 301

Qy        151 ProAlaValArgGlyLeuLeuGlySerCysValGlnLeuMetValValValGlyIleLeu 170
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        302 CCAGCAGTCCGGGGGTTGCTCGGCTCCTGTGTGCAGCTAATGGTCGTCGTCGGCATCCTC 361

Qy        171 LeuAlaTyrLeuAlaGlyTrpValLeuGluTrpArgTrpLeuAlaValLeuGlyCysVal 190
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        362 CTGGCCTACCTGGCAGGCTGGGTGCTGGAGTGGCGCTGGCTGGCTGTGCTGGGCTGCGTG 421

Qy        191 ProProSerLeuMetLeuLeuLeuMetCysPheMetProGluThrProArgPheLeuLeu 210
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        422 CCCCCCTCCCTCATGCTGCTTCTCATGTGCTTCATGCCCGAGACCCCGCGCTTCCTGCTG 481
```

```
Qy        211 ThrGlnHisArgArgGlnGluAlaMetAlaAlaLeuArgPheLeuTrpGlySer-GluGl 230
              |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db        482 ACTCAGCACAGGCGCCAGGAGGCCATGGCCGCCCTGCGGTTCCTGTGGGGCTCCTGAGCA 541

Qy        230 nGlyTrpGluAspProProIleGlyAlaGluGlnSerPheHisLeuAlaLeuLeuArgGl 250
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        542 GGGCTGGGAAGACCCCCCCATCGGGGCTGAGCAGAGCTTTCACCTGGCCCTGCTGCGGCA 601

Qy        250 nProGlyIleTyrLysProPheIleIleGlyValSerLeuMetAlaPheGlnGlnLeuSe 270
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        602 GCCCGGCATCTACAAGCCCTTCATCATCGGCGTCTCCCTGATGGCCTTCCAGCAGCTGTC 661

Qy        270 rGlyValAsnAlaValMetPheTyrAlaGluThrIlePheGluGluAlaLysPheLysAs 290
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        662 GGGGGTCAACGCCGTCATGTTCTATGCAGAGACCATCTTTGAAGAGGCCAAGTTCAAGGA 721

Qy        290 pSerSerLeuAla-SerValValValGlyValIleGlnValLeuPheThrAlaValAlaA 310
              ||||||||||||| ||||||||||||   |||||||||||||||||||||||||||   |
Db        722 CAGCAGCCTGGCCCTCGGTCGTCGTGCGTGTCATCCAGGTGCTGTTCACAGCTGTGGGGG 781

Qy        310 laLeuIleMetAspArgAlaGlyArgArgLeuLeuLeuValLeuSerGlyValValMetV 330
              |||||||||||||||||||||||   ||||||:::|||      |||||||||   ||||
Db        782 CTCTCATCATGGACAGAGCAGGCGG-AGGCTGGTCCTGGCTGG-TCAGGTGTGGCCATGG 839

Qy        330 alPheSerThrSerAlaPheGlyAlaTyrPheLysLeuThrGlnGlyGlyProGlyAsnS 350
              ||         ||||||   ||||||   ||||||||||||   |||   ||||||||||
Db        840 TGTCAGGCCGGAGTGCCTC-GGCGCCTAATTCAAGCTGACCCG-GGTTGGCCTGGCAATT 897

Qy        350 erSerHisValAlaIleSerAlaProValSerAla 361
              |||||   ||||||||||||      |||   |||
Db        898 CCTCGCC-GTGGCCATCTCGGGCCTTGTTCTTGCG 931

RESULT 8
BG700749
LOCUS       BG700749                 830 bp    mRNA    linear   EST 07-MAY-2001
DEFINITION  602681616F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:4814451 5',
            mRNA sequence.
ACCESSION   BG700749
VERSION     BG700749.1  GI:13970402
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 830)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
             cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10709  row: p  column: 04
            High quality sequence stop: 789.
FEATURES             Location/Qualifiers
     source          1..830
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4814451"
                     /tissue_type="hippocampus"
                     /lab_host="DH10B"
                     /clone_lib="NIH_MGC_95"
                     /note="Organ: brain; Vector: pBluescriptR (modified
                     pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI
                     (gtcgag); Oligo-dT primed using primer
```

```
                     5'-TTTTTTTTTTTTTTTTTVN-3', size-selected for  average
                     insert size 2.5 kb and normalized to ROT 5. This is a
                     primary library enriched for  full-length clones and
                     constructed using the Cap-trapper method (Carninci, in
                     preparation).  Library constructed by M. Brownstein
                     (NIMH/NHGRI, National Institutes of Health).  Note: this
                     is a NIH_MGC Library."
ORIGIN

Alignment Scores:
Pred. No.:              5e-107         Length:         830
Score:                  1293.00        Matches:        265
Percent Similarity:     96.01%         Conservative:   0
Best Local Similarity:  96.01%         Mismatches:     8
Query Match:            52.63%         Indels:         3
DB:                     12             Gaps:           0

US-09-886-954A-1 (1-477) x BG700749 (1-830)

Qy         133 AlaCysGlyValAlaSerLeuValAlaProValTyrIleSerGluIleAlaTyrProAla 152
               |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           5 GCCTGGGGTGTTGCCTCCCTAGTGGCCCCGGTCTACATCTCCGAAATCGCCTACCCAGCA 64

Qy         153 ValArgGlyLeuLeuGlySerCysValGlnLeuMetValValValGlyIleLeuLeuAla 172
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          65 GTCCGGGGGTTGCTCGGCTCCTGTGTGCAGCTAATGGTCGTCGTCGGCATCCTCCTGGCC 124

Qy         173 TyrLeuAlaGlyTrpValLeuGluTrpArgTrpLeuAlaValLeuGlyCysValProPro 192
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         125 TACCTGGCAGGCTGGGTGCTGGAGTGGCGCTGGCTGGCTGTGCTGGGCTGCGTGCCCCCC 184

Qy         193 SerLeuMetLeuLeuLeuMetCysPheMetProGluThrProArgPheLeuLeuThrGln 212
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         185 TCCCTCATGCTGCTTCTCATGTGCTTCATGCCCGAGACCCCGCGCTTCCTGCTGACTCAG 244

Qy         213 HisArgArgGlnGluAlaMetAlaAlaLeuArgPheLeuTrpGlySerGluGlnGlyTrp 232
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         245 CACAGGCGCCAGGAGGCCATGGCCGCCCTGCGGTTCCTGTGGGGCTCCGAGCAGGGCTGG 304

Qy         233 GluAspProProIleGlyAlaGluGlnSerPheHisLeuAlaLeuLeuArgGlnProGly 252
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         305 GAAGACCCCCCCATCGGGGCTGAGCAGAGCTTTCACCTGGCCCTGCTGCGGCAGCCCGGC 364

Qy         253 IleTyrLysProPheIleIleGlyValSerLeuMetAlaPheGlnGlnLeuSerGlyVal 272
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         365 ATCTACAAGCCCTTCATCATCGGCGTCTCCCTGATGGCCTTCCAGCAGCTGTCGGGGGTC 424

Qy         273 AsnAlaValMetPheTyrAlaGluThrIlePheGluGluAlaLysPheLysAspSerSer 292
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         425 AACGCCGTCATGTTCTATGCAGAGACCATCTTTGAAGAGGCCAAGTTCAAGGACAGCAGC 484

Qy         293 LeuAlaSerValValValGlyValIleGlnValLeuPheThrAlaValAlaAlaLeuIle 312
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         485 CTGGCCTCGGTCGTCGTGGGTGTCATCCAGGTGCTGTTCACAGCTGTGGCGGCTCTCATC 544

Qy         313 MetAspArgAlaGlyArgArgLeuLeuLeuValLeuSerGlyValValMetValPheSer 332
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         545 ATGGACAGAGCAGGGCGGAGGCTGCTCCTGGTCTTGTCAGGTGTGGTCATGGTGTTCAGC 604

Qy         333 ThrSerAlaPheGlyAlaTyrPheLysLeuThrGlnGlyGlyProGlyAsnSerSerHis 352
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         605 ACGAGTGCCTTCGGCGCCTACTTCAAGCTGACCCAGGGTGGCCCTGGCAACTCCTCGCAC 664

Qy         353 ValAlaIleSerAlaProValSerAla-GlnProVal-AspAlaSerValGlyLeuAlaT 372
               ||||||||||||||||||||||||||| ||||||||| ||||||||||||||||||||||
Db         665 GTGGCCATCTCGGCGCCTGTCTCTGCCACAGCCTGTTTGATGCCAGCGTGGGGCTGGCCT 724

Qy         372 rpLeu-AlaValGlySerMetCysLeuPheIleAlaGlyPheAlaValGlyTrpGlyPro 391
               ||||| ||||||||||||||||||||||||||||||            |||||||||
Db         725 GGCTGGGCCGTGGGCAGCATGTGCCTCTTCATCGCGCGTTTGCGGTGGGGCTGGGGGGCC 784

Qy         392 IleProTrpLeuLeuMetSerGluIlePheProLeuHisVal 405
               ||||||   |||||||||||||||||||||||||||   |||
```

```
Db         785 ATCCCCTTGCTCCTCATGTCAGAGATTTTCCCTCTTGCTGTC 826

RESULT 9
BI334832
LOCUS       BI334832                 991 bp    mRNA    linear   EST 30-JUL-2001
DEFINITION  602998959F1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:5141007 5',
            mRNA sequence.
ACCESSION   BI334832
VERSION     BI334832.1  GI:15019489
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 991)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: Incyte Genomics, Inc.
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11347  row: f  column: 16
            High quality sequence stop: 835.
FEATURES             Location/Qualifiers
     source          1. .991
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5141007"
                     /tissue_type="cervical carcinoma cell line"
                     /lab_host="DH10B"
                     /clone_lib="NIH_MGC_12"
                     /note="Organ: cervix; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.4 kb. Library prepared by Life
                     Technologies."
ORIGIN

Alignment Scores:
Pred. No.:              9.14e-105      Length:         991
Score:                  1269.50        Matches:        290
Percent Similarity:     65.05%         Conservative:   6
Best Local Similarity:  63.74%         Mismatches:     21
Query Match:            51.67%         Indels:         140
DB:                     12             Gaps:           6

US-09-886-954A-1 (1-477) x BI334832 (1-991)

Qy           1 MetThrProGluAspProGluGluThrGlnProLeuLeuGlyProProGlyGlySerAla 20
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          17 ATGACGCCCGAGGACCCAGAGGAAACCCAGCCGCTTCTGGGGCCTCCTGGCGGCAG---- 72

Qy          21 ProArgGlyArgArgValPheLeuAlaAlaPheAlaAlaAlaLeuGlyProLeuSerPhe 40

Db          72 ------------------------------------------------------------ 72

Qy          41 GlyPheAlaLeuGlyTyrSerSerProAlaIleProSerLeuGlnArgAlaAlaProPro 60

Db          72 ------------------------------------------------------------ 72

Qy          61 AlaProArgLeuAspAspAlaAlaAlaSerTrpPheGlyAlaValValThrLeuGlyAla 80

Db          72 ------------------------------------------------------------ 72

Qy          81 AlaAlaGlyGlyValLeuGlyGlyTrpLeuValAspArgAlaGlyArgLysLeuSerLeu 100

Db          72 ------------------------------------------------------------ 72
```

```
Qy        101 LeuLeuCysSerValProPheValAlaGlyPheAlaValIleThrAlaAlaGlnAspVal 120

Db         72 ------------------------------------------------------------ 72

Qy        121 TrpMetLeuLeuGlyGlyArgLeuLeuThrGlyLeuAlaCysGlyValAlaSerLeuVal 140

Db         72 ------------------------------------------------------------ 72

Qy        141 AlaProValTyrIleSerGluIleAlaTyrProAlaValArgGlyLeuLeuGlySerCys 160
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         73 ------GTCTACATCTCCGAAATCGCCTACCCAGCAGTCCGGGGGTTGCTCGGCTCCTGT 126

Qy        161 ValGlnLeuMetValValValGlyIleLeuLeuAlaTyrLeuAlaGlyTrpValLeuGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        127 GTGCAGCTAATGGTCGTCGTCGGCATCCTCCTGGCCTACCTGGCAGGCTGGGTGCTGGAG 186

Qy        181 TrpArgTrpLeuAlaValLeuGlyCysValProProSerLeuMetLeuLeuLeuMetCys 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        187 TGGCGCTGGCTGGCTGTGCTGGGCTGCGTGCCCCCCTCCCTCATGCTGCTTCTCATGTGC 246

Qy        201 PheMetProGluThrProArgPheLeuLeuThrGlnHisArgArgGlnGluAlaMetAla 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        247 TTCATGCCCGAGACCCCGCGCTTCCTGCTGACTCAGCACAGGCGCCAGGAGGCCATGGCC 306

Qy        221 AlaLeuArgPheLeuTrpGlySerGluGlnGlyTrpGluAspProProIleGlyAlaGlu 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        307 GCCCTGCGGTTCCTGTGGGGCTCCGAGCAGGGCTGGGAAGACCCCCCCATCGGGGCTGAG 366

Qy        241 GlnSerPheHisLeuAlaLeuLeuArgGlnProGlyIleTyrLysProPheIleIleGly 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        367 CAGAGCTTTCACCTGGCCCTGCTGCGGCAGCCCGGCATCTACAAGCCCTTCATCATCGGC 426

Qy        261 ValSerLeuMetAlaPheGlnGlnLeuSerGlyValAsnAlaValMetPheTyrAlaGlu 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        427 GTCTCCCTGATGGCCTTCCAGCAGCTGTCGGGGGTCAACGCCGTCATGTTCTATGCAGAG 486

Qy        281 ThrIlePheGluGluAlaLysPheLysAspSerSerLeuAlaSerValValValGlyVal 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        487 ACCATCTTTGAAGAGGCCAAGTTCAAGGACAGCAGCCTGGCCTCGGTCGTCGTGGGTGTC 546

Qy        301 IleGlnValLeuPheThrAlaValAlaAlaLeuIleMetAspArgAlaGlyArgArgLeu 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        547 ATCCAGGTGCTGTTCACAGCTGTGGCGGCTCTCATCATGGACAGAGCAGGGCGGAGGCTG 606

Qy        321 LeuLeuValLeuSerGlyValValMetValPheSerThrSerAlaPheGlyAlaTyrPhe 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        607 CTCCTGGTCTTGTCAGGTGTGGTCATGGTGTTCAGCACGAGTGCCTTCGGCGCCTACTTC 666

Qy        341 LysLeuThrGlnGlyGlyProGlyAsnSerSer-HisValAlaIleSerAlaProValSe 360
              ||||||||||||||||||||||||||||||||| :::|||||||||||||||||||||||
Db        667 AAGCTGACCCAGGGTGGCCCTGGCAACTCCTCGCAACGTGGCCATCTCGGCGCCTGTCTC 726

Qy        360 rAlaGln-ProValAspAlaSerVal-GlyLeuAlaTrpLeuAlaVal-GlySerMetCy 379
              ||||||| |||||||||||||||||| ||||||||||||||||||||| |||||||||||
Db        727 TGCACAAGCCTGTTGATGCCAGCGTGGGGCCTGGCCTGGCTGGCCGTTGGGCAGCATGTG 786

Qy        379 sLeuPheIleAlaGlyPheAlaValGly---TrpGlyProIleProTrpLeu---LeuMe 397
              ||||||||||||||||||||||||||||      ||||||   ||||||:::
Db        787 CCTCTTCATCGCCGGCTTTGCGGTGGGCTTGGGGGGCCCATTCCCCTGGGTCCCTCATTG 846

Qy        397 tSerGluIlePheProLeuHisValLysGlyValAlaThrGly---------------Il 412
                  :::||||||||||||            :::   ||||||
Db        847 TCAGAAGATCTTTCCCCTC---------TGCATTGTCACGGGCCGTGGCGACAGGGCATT 897

Qy        412 eCysValLeuThrAsnTrpLeuMetAlaPheLeuValThrLysGluPheSerSerLeuMe 432
                  |||   ||||||      |||||||||      |||   |||||||||   |||::
Db        898 CTAGGTCCCCACAAATGGCTCCATGGCCTTTCCCGGGACCAG-GAGTTCAGCCAGCTCTT 956

Qy        432 tGluValLeuArgProTyrGlyAlaPheTrpLeuAla 444
              :||||||   ||||||            |||:::|||
Db        957 GGAAGTCCCCAGGCCCTTGAAC---CTCTGGGTTGCT 990
```

```
RESULT 10
AY414181
LOCUS       AY414181                1004 bp    DNA     linear   GSS 17-DEC-2003
DEFINITION  Pan troglodytes SLC2A8 gene, VIRTUAL TRANSCRIPT, partial sequence,
            genomic survey sequence.
ACCESSION   AY414181
VERSION     AY414181.1  GI:39770143
KEYWORDS    GSS.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (bases 1 to 1004)
  AUTHORS   Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
            Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
            Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
            Adams,M.D. and Cargill,M.
  TITLE     Inferring nonneutral evolution from human-chimp-mouse orthologous
            gene trios
  JOURNAL   Science 302 (5652), 1960-1963 (2003)
   PUBMED   14671302
REFERENCE   2  (bases 1 to 1004)
  AUTHORS   Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
            Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
            Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
            Adams,M.D. and Cargill,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
            Rockville, MD 20850, USA
COMMENT     This sequence was made by sequencing genomic exons and ordering
            them based on alignment.
FEATURES             Location/Qualifiers
     source          1..1004
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
     gene            <1..>1004
                     /gene="SLC2A8"
                     /locus_tag="HCM5138"
ORIGIN

Alignment Scores:
Pred. No.:              1.04e-104      Length:          1004
Score:                  1269.00        Matches:         263
Percent Similarity:     79.88%         Conservative:    3
Best Local Similarity:  78.98%         Mismatches:      67
Query Match:            51.65%         Indels:          0
DB:                     29             Gaps:            0

US-09-886-954A-1 (1-477) x AY414181 (1-1004)

Qy        145 IleSerGluIleAlaTyrProAlaValArgGlyLeuLeuGlySerCysValGlnLeuMet 164
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          3 ATCTCCGAAATCGCCTACCCAGCAGTCCGGGGGTTGCTCGGCTCTTGTGTGCAGCTGATG 62

Qy        165 ValValValGlyIleLeuLeuAlaTyrLeuAlaGlyTrpValLeuGluTrpArgTrpLeu 184
              |||||||||||||||||||||||||||||||||
Db         63 GTCGTCGTCGGCATCCTCCTGGCCTACCTGGCAGNNNNNNNNNNNNNNNNNNNNNNNNNN 122

Qy        185 AlaValLeuGlyCysValProProSerLeuMetLeuLeuLeuMetCysPheMetProGlu 204

Db        123 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 182

Qy        205 ThrProArgPheLeuLeuThrGlnHisArgArgGlnGluAlaMetAlaAlaLeuArgPhe 224

Db        183 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 242

Qy        225 LeuTrpGlySerGluGlnGlyTrpGluAspProProIleGlyAlaGluGlnSerPheHis 244
                                                                 |||||||||
Db        243 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNAGCTTTCAC 302

Qy        245 LeuAlaLeuLeuArgGlnProGlyIleTyrLysProPheIleIleGlyValSerLeuMet 264
              |||||||||||||||||||||||||||||||||||||||:::||||||||||||||||||
```

```
Db        303 CTGGCCCTGCTGCGGCAGCCCGGCATCTACAAGCCCTTCGTCATCGGCGTCTCCCTGATG 362

Qy        265 AlaPheGlnGlnLeuSerGlyValAsnAlaValMetPheTyrAlaGluThrIlePheGlu 284
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        363 GCCTTCCAGCAGCTGTCGGGGGTCAACGCCGTCATGTTCTATGCAGAGACCATCTTTGAA 422

Qy        285 GluAlaLysPheLysAspSerSerLeuAlaSerValValValGlyValIleGlnValLeu 304
              ||||||||||||||||||||||||||||||||||||:::|||||||||||||||||||||
Db        423 GAGGCCAAGTTCAAGGACAGCAGCCTGGCCTCGGTCATCGTGGGTGTCATCCAGGTGCTG 482

Qy        305 PheThrAlaValAlaAlaLeuIleMetAspArgAlaGlyArgArgLeuLeuLeuValLeu 324
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        483 TTCACAGCTGTGGCGGCTCTCATCATGGACAGAGCAGGGCGGAGGCTGCTCCTGGTCNNG 542

Qy        325 SerGlyValValMetValPheSerThrSerAlaPheGlyAlaTyrPheLysLeuThrGln 344
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        543 TCAGGTGTGGTCATGGTGTTCAGCACGAGTGCCTTCGGCGCCTACTTCAAGCTGACCCAG 602

Qy        345 GlyGlyProGlyAsnSerSerHisValAlaIleSerAlaProValSerAlaGlnProVal 364
              ||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||
Db        603 GGTGGCCCTGGCAACTCCTCGCACGTGGCCCTCTCGGCGCCTGTCTCTGCACAGCCTGTT 662

Qy        365 AspAlaSerValGlyLeuAlaTrpLeuAlaValGlySerMetCysLeuPheIleAlaGly 384
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        663 GATGCCAGCGTGGGGCTGGCCTGGCTGGCCGTGGGCAGCATGTGCCTCTTCATCGCCGGC 722

Qy        385 PheAlaValGlyTrpGlyProIleProTrpLeuLeuMetSerGluIlePheProLeuHis 404
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        723 TTTGCGGTGGGCTGGGGGCCCATCCCCTGGCTCCTCATGTCAGAGATCTTCCCTCTGCAT 782

Qy        405 ValLysGlyValAlaThrGlyIleCysValLeuThrAsnTrpLeuMetAlaPheLeuVal 424
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        783 GTCAAGGGCGTGGCGACAGGCATCTGCGTCCTCACCAACTGGCTCATGGCCTTTCTCGTG 842

Qy        425 ThrLysGluPheSerSerLeuMetGluValLeuArgProTyrGlyAlaPheTrpLeuAla 444
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        843 ACCAAGGAGTTCAGCAGCCTCATGGAGGTCCTCAGGCCCTATGGAGCCTTCTGGCTTGCC 902

Qy        445 SerAlaPheCysIlePheSerValLeuPheThrLeuPheCysValProGluThrLysGly 464
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        903 TCCGCTTTCTGCATCTTCAGTGTCCTTTTCACTTTGTTCTGTGTCCCTGAAACTAAAGGA 962

Qy        465 LysThrLeuGluGlnIleThrAlaHisPheGluGlyArg 477
              |||||||||||||||||||||||||||||||||||||||
Db        963 AAGACTCTGGAACAAATCACAGCCCATTTTGAGGGGCGA 1001

RESULT 11
CD557801
LOCUS       CD557801                 875 bp    mRNA    linear   EST 11-JUN-2003
DEFINITION  AGENCOURT_14413291 NIH_MGC_180 Homo sapiens cDNA clone
            IMAGE:30386912 5', mRNA sequence.
ACCESSION   CD557801
VERSION     CD557801.1  GI:31583869
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 875)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Daniela S. Gerhard, Ph.D.
            Office of Cancer Genomics
            National Cancer Institute / NIH
            Bldg. 31 Rm10A07 Bethesda, MD 20892
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Michael Brownstein
             cDNA Library Preparation: Invitrogen Corp
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: NDAM456  row: f  column: 09
            High quality sequence stop: 618.
FEATURES             Location/Qualifiers
     source          1..875
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:30386912"
                     /lab_host="DH10B-Ton A ( T1 and T5 phage resistances)"
                     /clone_lib="NIH_MGC_180"
                     /note="Organ: Testis; Vector: pCMV-SPORT6.1; Site_1: NotI;
                     Site_2: EcoRV (destroyed); Library is oligo-dT primed and
                     directionally cloned (EcoRV site is destroyed upon
                     cloning). Average insert size 1.68 kb. Library was
                     constructed by (Invitrogen). Note: this is a NIH_MGC
                     Library."
ORIGIN

Alignment Scores:
Pred. No.:              1.26e-104     Length:         875
Score:                  1267.00       Matches:        265
Percent Similarity:     95.05%        Conservative:   4
Best Local Similarity:  93.64%        Mismatches:     7
Query Match:            51.57%        Indels:         7
DB:                     14            Gaps:           0

US-09-886-954A-1 (1-477) x CD557801 (1-875)

Qy         76 ValThrLeuGlyAlaAlaAlaGlyGlyValLeuGlyGlyTrpLeuValAspArgAlaGly 95
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTGACCCTGGGTGCCGCGGCGGGGGGAGTGCTGGGCGGCTGGCTGGTGGACCGCGCCGGG 60

Qy         96 ArgLysLeuSerLeuLeuLeuCysSerValProPheValAlaGlyPheAlaValIleThr 115
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCAAGCTGAGCCTCTTGCTGTGCTCCGTGCCCTTCGTGGCCGGCTTTGCCGTCATCACC 120

Qy        116 AlaAlaGlnAspValTrpMetLeuLeuGlyGlyArgLeuLeuThrGlyLeuAlaCysGly 135
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCGGCCCAGGACGTGTGGATGCTGCTGGGGGGCCGCCTCCTCACCGGCCTGGCCTGCGGT 180

Qy        136 ValAlaSerLeuValAlaProValTyrIleSerGluIleAlaTyrProAlaValArgGly 155
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GTTGCCTCCCTAGTGGCCCCGGTCTACATCTCCGAAATCGCCTACCCAGCAGTCCGGGGG 240

Qy        156 LeuLeuGlySerCysValGlnLeuMetValValValGlyIleLeuLeuAlaTyrLeuAla 175
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTGCTCGGCTCCTGTGTGCAGCTAATGGTCGTCGTCGGCATCCTCCTGGCCTACCTGGCA 300

Qy        176 GlyTrpValLeuGluTrpArgTrpLeuAlaValLeuGlyCysValProProSerLeuMet 195
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GGCTGGGTGCTGGAGTGGCGCTGGCTGGCTGTGCTGGGCTGCGTGCCCCCCTCCCTCATG 360

Qy        196 LeuLeuLeuMetCysPheMetProGluThrProArgPheLeuLeuThrGlnHisArgArg 215
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CTGCTTCTCATGTGCTTCATGCCCGAGACCCCGCGCTTCCTGCTGACTCAGCACAGGCGC 420

Qy        216 GlnGluAlaMetAlaAlaLeuArgPheLeuTrpGlySerGluGlnGlyTrpGluAspPro 235
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CAGGAGGCCATGGCCGCCCTGCGGTTCCTGTGGGGCTCCGAGCAGGGCTGGGAAGACCCC 480

Qy        236 ProIleGlyAlaGluGlnSerPheHisLeuAlaLeuLeuArgGlnProGlyIleTyrLys 255
              |||||||||||||||||||||||||||||||||||||||||||||||||||   ||||||
Db        481 CCCATCGGGGCTGAGCAGAGCTTTCACCTGGCCCTGCTGCGGCAGCCCGGCACCTACAAG 540

Qy        256 ProPheIleIleGlyValSerLeuMetAlaPheGlnGlnLeuSerGlyValAsnAlaVal 275
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CCCTTCATCATCGGCGTCTCCCTGATGGCCTTCCAGCAGCTGTCGGGGGTCAACGCCGTC 600

Qy        276 MetPheTyrAlaGluThrIlePheGluGluAlaLysPheLysAspSerSerLeuAlaSer 295
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 ATGTTCTATGCAGAGACCATCTTTGAAGAGGCCAAGTTCAAGGACAGCAGCCTGGCCTCG 660
```

```
Qy        296 ValValValGlyValIleGlnValLeuPheThrAlaValAlaAlaLeuIleMetAspArg 315
              |||||||||||||||||||||||||||||||||||||||||||||   :::||||||:::
Db        661 GTCGTCGTGGGTGTCATCCAGGTGCTGTTCACAGCTGTGGCGGCTCCTCTCATGGACAAA 720

Qy        316 AlaGlyArgArgLeuLeuLeu-ValLeuSerGlyVal-ValMet-ValPheSerThrSer 334
                 |||||||||||||||||| ||||||||||||||| |||||| ||||||||||||:::
Db        721 CCAGGGCGGAGGCTGCTCCTGGGTCTTGTCAGGGGTGGGTCATGGGTGTTCAGCACAAAT 780

Qy        335 Ala--PheGlyAlaTyrPheLysLeu--ThrGlnGlyGlyProGlyAsnSerSer 351
              |||  ||||||   |||||||||     |||||||||||||||      :::|||
Db        781 GCCCTTTCGGCCCCTACTTCAAAGCTGAACCCAGGGTGGCCCCGGGCAAATTCC 835

RESULT 12
BM044230
LOCUS       BM044230                 880 bp    mRNA    linear   EST 07-NOV-2001
DEFINITION  603621486F1 NIH_MGC_40 Homo sapiens cDNA clone IMAGE:5447181 5',
            mRNA sequence.
ACCESSION   BM044230
VERSION     BM044230.1  GI:16773497
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 880)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1930  row: k  column: 22
            High quality sequence stop: 869.
FEATURES             Location/Qualifiers
     source          1..880
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5447181"
                     /tissue_type="carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_40"
                     /note="Organ: prostate; Vector: pOTB7; Site_1: XhoI;
                     Site_2: EcoRI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G). Library constructed  by
                     Ling Hong in the laboratory of Gerald M. Rubin (University
                     of California,  Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life  Technologies).
                     Note: this is a NIH_MGC Library."
ORIGIN

Alignment Scores:
Pred. No.:                2.67e-102      Length:          880
Score:                    1241.50        Matches:         280
Percent Similarity:       58.82%         Conservative:    0
Best Local Similarity:    58.82%         Mismatches:      5
Query Match:              50.53%         Indels:          193
DB:                       12             Gaps:            2

US-09-886-954A-1 (1-477) x BM044230 (1-880)

Qy          4 GluAspProGluGluThrGlnProLeuLeuGlyProProGlyGlySerAlaProArgGly 23
              |||||||||||||||||||||||||||||||||||||||||||||
Db          4 GAGGACCCAGAGGAAACCCAGCCGCTTCTGGGGCCTCCTGGCGGCAG------------- 50

Qy         24 ArgArgValPheLeuAlaAlaPheAlaAlaAlaLeuGlyProLeuSerPheGlyPheAla 43

Db         50 ------------------------------------------------------------ 50

Qy         44 LeuGlyTyrSerSerProAlaIleProSerLeuGlnArgAlaAlaProProAlaProArg 63

Db         50 ------------------------------------------------------------ 50

Qy         64 LeuAspAspAlaAlaAlaSerTrpPheGlyAlaValValThrLeuGlyAlaAlaAlaGly 83

Db         50 ------------------------------------------------------------ 50

Qy         84 GlyValLeuGlyGlyTrpLeuValAspArgAlaGlyArgLysLeuSerLeuLeuLeuCys 103

Db         50 ------------------------------------------------------------ 50

Qy        104 SerValProPheValAlaGlyPheAlaValIleThrAlaAlaGlnAspValTrpMetLeu 123

Db         50 ------------------------------------------------------------ 50

Qy        124 LeuGlyGlyArgLeuLeuThrGlyLeuAlaCysGlyValAlaSerLeuValAlaProVal 143
                                                                       |||
Db         51 ---------------------------------------------------------GTC 53

Qy        144 TyrIleSerGluIleAlaTyrProAlaValArgGlyLeuLeuGlySerCysValGlnLeu 163
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         54 TACATCTCCGAAATCGCCTACCCAGCAGTCCGGGGGTTGCTCGGCTCCTGTGTGCAGCTA 113

Qy        164 MetValValValGlyIleLeuLeuAlaTyrLeuAlaGlyTrpValLeuGluTrpArgTrp 183
              ||||||||||||||||||||||||||||||||||||
Db        114 ATGGTCGTCGTCGGCATCCTCCTGGCCTACCTGGCA------------------------ 149

Qy        184 LeuAlaValLeuGlyCysValProProSerLeuMetLeuLeuLeuMetCysPheMetPro 203

Db        149 ------------------------------------------------------------ 149

Qy        204 GluThrProArgPheLeuLeuThrGlnHisArgArgGlnGluAlaMetAlaAlaLeuArg 223

Db        149 ------------------------------------------------------------ 149

Qy        224 PheLeuTrpGlySerGluGlnGlyTrpGluAspProProIleGlyAlaGluGlnSer-Ph 243
                                                                        ||
Db        150 ------------------------------------------------------GAGCTT 155

Qy        243 eHisLeuAlaLeuLeuArgGlnProGlyIleTyrLysProPheIleIleGlyValSerLe 263
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        156 TCACCTGGCCCTGCTGCGGCAGCCCGGCATCTACAAGCCCTTCATCATCGGCGTCTCCCT 215

Qy        263 uMetAlaPheGlnGlnLeuSerGlyValAsnAlaValMetPheTyrAlaGluThrIlePh 283
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        216 GATGGCCTTCCAGCAGCTGTCGGGGGTCAACGCCGTCATGTTCTATGCAGAGACCATCTT 275

Qy        283 eGluGluAlaLysPheLysAspSerSerLeuAlaSerValValValGlyValIleGlnVa 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        276 TGAAGAGGCCAAGTTCAAGGACAGCAGCCTGGCCTCGGTCGTCGTGGGTGTCATCCAGGT 335

Qy        303 lLeuPheThrAlaValAlaAlaLeuIleMetAspArgAlaGlyArgArgLeuLeuLeuVa 323
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        336 GCTGTTCACAGCTGTGGCGGCTCTCATCATGGACAGAGCAGGGCGGAGGCTGCTCCTGGT 395

Qy        323 lLeuSerGlyValValMetValPheSerThrSerAlaPheGlyAlaTyrPheLysLeuTh 343
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        396 CTTGTCAGGTGTGGTCATGGTGTTCAGCACGAGTGCCTTCGGCGCCTACTTCAAGCTGAC 455

Qy        343 rGlnGlyGlyProGlyAsnSerSerHisValAlaIleSerAlaProValSerAlaGlnPr 363
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        456 CCAGGGTGGCCCTGGCAACTCCTCGCACGTGGCCATCTCGGCGCCTGTCTCTGCACAGCC 515

Qy        363 oValAspAlaSerValGlyLeuAlaTrpLeuAlaValGlySerMetCysLeuPheIleAl 383
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        516 TGTTGATGCCAGCGTGGGGCTGGCCTGGCTGGCCGTGGGCAGCATGTGCCTCTTCATCGC 575

Qy        383 aGlyPheAlaValGlyTrpGlyProIleProTrpLeuLeuMetSerGluIlePheProLe 403
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       576 CGGCTTTGCGGTGGGCTGGGGGCCCATCCCCTGGCTCCTCATGTCAGAGATCTTCCCTCT 635

Qy       403 uHisValLysGlyValAlaThrGlyIleCysValLeuThrAsnTrpLeuMetAlaPheLe 423
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       636 GCATGTCAAGGGCGTGGCGACAGGCATCTGCGTCCTCACCAACTGGCTCATGGCCTTTCT 695

Qy       423 uValThrLysGluPheSerSerLeuMetGluValLeuArgProTyrGlyAlaPheTrpLe 443
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       696 CGTGACCAAGGAGTTCAGCAGCCTCATGGAGGTCCTCAGGCCCTATGGAGCCTTCTGGCT 755

Qy       443 uAlaSerAlaPheCysIlePheSerValLeuPheThrLeuPheCys-ValProGluThrL 463
             |||||||   |||||||||||||||||||||||||||||||||    |||||||||||||
Db       756 TGCCTCCGTTTCTGCATCTTCAGTGTCCTTTTCACTTTGTTCTGGTGTCCCTGAAACTA 815

Qy       463 ysGlyLysThrLeuGluGlnIleThrAlaHisPheGluGlyArg 477
             |||||||||||||||||||||||||||||   ||||||||||||
Db       816 AAGGAAAGACTCTGGAACAAATCACAGCCA--TTTGAGGGGCGA 857

RESULT 13
BG478000
LOCUS       BG478000                 828 bp    mRNA    linear   EST 21-MAR-2001
DEFINITION  602522606F1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:4641145 5',
            mRNA sequence.
ACCESSION   BG478000
VERSION     BG478000.1  GI:13410279
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 828)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC/DCTD/DTP
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1407  row: k  column: 02
            High quality sequence stop: 788.
FEATURES             Location/Qualifiers
     source          1..828
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4641145"
                     /tissue_type="melanotic melanoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_20"
                     /note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Size-selected >500bp for average
                     insert size 1.8kb.  Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using  ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
ORIGIN

Alignment Scores:
Pred. No.:              5.04e-102      Length:         828
Score:                  1238.00        Matches:        250
Percent Similarity:     95.08%         Conservative:   1
Best Local Similarity:  94.70%         Mismatches:     8
Query Match:            50.39%         Indels:         5
DB:                     12             Gaps:           0

US-09-886-954A-1 (1-477) x BG478000 (1-828)
```

```
Qy         1 MetThrProGluAspProGluGluThrGlnProLeuLeuGlyProProGlyGlySerAla 20
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        47 ATGACGCCCGAGGACCCAGAGGAAACCCAGCCGCTTCTGGGGCCTCCTGGCGGCAGCGCG 106

Qy        21 ProArgGlyArgArgValPheLeuAlaAlaPheAlaAlaAlaLeuGlyProLeuSerPhe 40
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       107 CCCCGCGGCCGCCGCGTCTTCCTCGCCGCCTTCGCCGCTGCCCTGGGCCCACTCAGCTTC 166

Qy        41 GlyPheAlaLeuGlyTyrSerSerProAlaIleProSerLeuGlnArgAlaAlaProPro 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       167 GGCTTCGCGCTCGGCTACAGCTCCCCGGCCATCCCTAGCCTGCAGCGCGCCGCGCCCCCG 226

Qy        61 AlaProArgLeuAspAspAlaAlaAlaSerTrpPheGlyAlaValValThrLeuGlyAla 80
             ||||||||||||   |||||||||||||||||||||||||||||||||||||||||||||
Db       227 GCCCCGCGCCTGCACGACGCCGCCGCCTCCTGGTTCGGGGCTGTCGTGACCCTGGGTGCC 286

Qy        81 AlaAlaGlyGlyValLeuGlyGlyTrpLeuValAspArgAlaGlyArgLysLeuSerLeu 100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       287 GCGGCGGGGGGAGTGCTGGGCGGCTGGCTGGTGGACCGCGCCGGGCGCAAGCTGAGCCTC 346

Qy       101 LeuLeuCysSerValProPheValAlaGlyPheAlaValIleThrAlaAlaGlnAspVal 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       347 TTGCTGTGCTCCGTGCCCTTCGTGGCCGGCTTTGCCGTCATCACCGCGGCCCAGGACGTG 406

Qy       121 TrpMetLeuLeuGlyGlyArgLeuLeuThrGlyLeuAlaCysGlyValAlaSerLeuVal 140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       407 TGGATGCTGCTGGGGGGCCGCCTCCTCACCGGCCTGGCCTGCGGTGTTGCCTCCCTAGTG 466

Qy       141 AlaProValTyrIleSerGluIleAlaTyrProAlaValArgGlyLeuLeuGlySerCys 160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       467 GCCCCGGTCTACATCTCCGAAATCGCCTACCCAGCAGTCCGGGGGTTGCTCGGCTCCTGT 526

Qy       161 ValGlnLeuMetValValValGlyIleLeuLeuAla--TyrLeuAlaGlyTrpValLeuG 180
             |||||||||||||||||||||||||||||||||     ||||||||||||||||||||||
Db       527 GTGCAGCTAATGGTCGTCGTCGGCATCCTCCTTGGCCTTACCTGGCAGGCTGGGTGCTGG 586

Qy       180 luTrpArgTrpLeuAlaValLeuGlyCysVal-ProProSerLeuMetLeuLeuLeuMet 199
             |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db       587 AGTGGCGCTGGCTGGCTGTGCTGGGCTGCGTTGCCCCCTTCCCTCATGCTGCTTCTCATG 646

Qy       200 CysPheMetProGluThrProArgPheLeuLeuThrGlnHisArgArgGlnGluAlaMet 219
             |||||||||||||||||||||   ||||||||||||||||||||||||||||||||||||
Db       647 TGCTTCATGCCCGAGACCCCGAGCTTCCTGCTGACTCAGCACAGGCGCCAGGAGGCCATG 706

Qy       220 AlaAlaLeuArgPheLeuTrpGlySerGluGlnGlyTrpGluAspProProIleGlyAla 239
             ||||||         |||   |||||||||||||||||||||||||||||||||||||||
Db       707 GCCGCCTGCGGTTCCTTGTGAGGCTCCGAGCAGGGCTGGGAAGACCCCCCCATCGGGGCT 766

Qy       240 GluGlnSerPheHisLeuAla--LeuLeuArgGlnProGlyIleTyrLysProPheIleI 259
             ||||||||||||||||||     |||||||||:::|||||||||||||||||||||||||
Db       767 GAGCAGAGCTTTCACCTTGGCCCTTGTTGCGGAAGCCCGGCATCTACAAGCCCTTCATCA 826

Qy       259 le 259
             ||
Db       827 TC 828

RESULT 14
BG717034
LOCUS       BG717034                 777 bp    mRNA    linear   EST 08-MAY-2001
DEFINITION  602689093F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:4821598 5',
            mRNA sequence.
ACCESSION   BG717034
VERSION     BG717034.1  GI:13996221
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 777)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
```

```
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
             cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10728  row: i  column: 23
            High quality sequence stop: 722.
FEATURES             Location/Qualifiers
     source          1..777
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4821598"
                     /lab_host="DH10B"
                     /clone_lib="NIH_MGC_97"
                     /note="Organ: testis; Vector: pBluescriptR (modified
                     pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI
                     (gtcgag); Oligo-dT primed using primer
                     5'-TTTTTTTTTTTTTTTTVN-3', size-selected for  average
                     insert size 2.2 kb and normalized to ROT 5. This is a
                     primary library enriched for full-length clones and
                     constructed using the Cap-trapper method (Carninci, in
                     preparation).  Library constructed by M. Brownstein
                     (NIMH/NHGRI, National Institutes of Health). Note: this is
                     a NIH_MGC Library."
ORIGIN

Alignment Scores:
Pred. No.:                1.3e-101      Length:         777
Score:                    1233.00       Matches:        243
Percent Similarity:       98.39%        Conservative:   1
Best Local Similarity:    97.98%        Mismatches:     3
Query Match:              50.18%        Indels:         2
DB:                       12            Gaps:           0

US-09-886-954A-1 (1-477) x BG717034 (1-777)

Qy          1 MetThrProGluAspProGluGluThrGlnProLeuLeuGlyProProGlyGlySerAla 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         36 ATGACGCCCGAGGACCCAGAGGAAACCCAGCCGCTTCTGGGGCCTCCTGGCGGCAGCGCG 95

Qy         21 ProArgGlyArgArgValPheLeuAlaAlaPheAlaAlaAlaLeuGlyProLeuSerPhe 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         96 CCCCGCGGCCGCCGCGTCTTCCTCGCCGCCTTCGCCGCTGCCCTGGGCCCACTCAGCTTC 155

Qy         41 GlyPheAlaLeuGlyTyrSerSerProAlaIleProSerLeuGlnArgAlaAlaProPro 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        156 GGCTTCGCGCTCGGCTACAGCTCCCCGGCCATCCCTAGCCTGCAGCGCGCCGCGCCCCCG 215

Qy         61 AlaProArgLeuAspAspAlaAlaAlaSerTrpPheGlyAlaValValThrLeuGlyAla 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        216 GCCCCGCGCCTGGACGACGCCGCCGCCTCCTGGTTCGGGGCTGTCGTGACCCTGGGTGCC 275

Qy         81 AlaAlaGlyGlyValLeuGlyGlyTrpLeuValAspArgAlaGlyArgLysLeuSerLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        276 GCGGCGGGGGGAGTGCTGGGCGGCTGGCTGGTGGACCGCGCCGGGCGCAAGCTGAGCCTC 335

Qy        101 LeuLeuCysSerValProPheValAlaGlyPheAlaValIleThrAlaAlaGlnAspVal 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        336 TTGCTGTGCTCCGTGCCCTTCGTGGCCGGCTTTGCCGTCATCACCGCGGCCCAGGACGTG 395

Qy        121 TrpMetLeuLeuGlyGlyArgLeuLeuThrGlyLeuAlaCysGlyValAlaSerLeuVal 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        396 TGGATGCTGCTGGGGGGCCGCCTCCTCACCGGCCTGGCCTGCGGTGTTGCCTCCCTAGTG 455

Qy        141 AlaProValTyrIleSerGluIleAlaTyrProAlaValArgGlyLeuLeuGlySerCys 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        456 GCCCCGGTCTACATCTCCGAAATCGCCTACCCAGCAGTCCGGGGGTTGCTCGGCTCCTGT 515
```

```
Qy        161 ValGlnLeuMetValValValGlyIleLeuLeuAlaTyrLeuAlaGlyTrpValLeuGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        516 GTGCAGCTAATGGTCGTCGTCGGCATCCTCCTGGCCTACCTGGCAGGCTGGGTGCTGGAG 575

Qy        181 TrpArgTrpLeuAlaValLeuGlyCysValProProSerLeuMetLeuLeuLeuMetCys 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        576 TGGCGCTGGCTGGCTGTGCTGGGCTGCGTGCCCCCCTCCCTCATGCTGCTTCTCATGTGC 635

Qy        201 PheMetProGluThrProArgPheLeuLeuThrGlnHisArgArgGlnGluAlaMetAla 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        636 TTCATGCCCGAGACCCCGCGCTTCCTGCTGACTCAGCACAGGCGCCAGGAGGCCATGGCC 695

Qy        221 AlaLeuArgPheLeu-TrpGlySerGluGlnGlyTrpGluAspProProIleGlyAlaGl 240
              ||||||||||||||| ||||||||||||   |||||||||||||||||    ||||||::
Db        696 GCCCTGCGGTTCCTGTTGGGGCTCCGAGCATGGCTGGGAAGACCCCCCATC-GGAGCTGA 754

Qy        240 uGlnSerPheHisLeuAlaLeu 247
              :|||||||||||||||   |||
Db        755 CCAGAGCTTCCACCTGGACCTG 776

RESULT 15
BI757409
LOCUS       BI757409                 889 bp    mRNA    linear   EST 25-SEP-2001
DEFINITION  603029344F1 NIH_MGC_114 Homo sapiens cDNA clone IMAGE:5199778 5',
            mRNA sequence.
ACCESSION   BI757409
VERSION     BI757409.1  GI:15748987
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 889)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11500  row: g  column: 11
            High quality sequence stop: 877.
FEATURES             Location/Qualifiers
     source          1..889
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5199778"
                     /lab_host="DH10B"
                     /clone_lib="NIH_MGC_114"
                     /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: EcoRV (destroyed); RNA source anonymous pool of 6
                     male brains, age range 23-27 yo. Library is oligo-dT
                     primed and directionally cloned (EcoRV site is  destroyed
                     upon cloning).  Average insert size 1.5 kb, insert size
                     range 1-3 kb.  Library is normalized and enriched for
                     full-length  clones and was constructed by C. Gruber
                     (Invitrogen).  Research  Genetics tracking code 019. Note:
                     this is a NIH_MGC Library."
ORIGIN

Alignment Scores:
Pred. No.:                2.72e-101     Length:         889
Score:                    1230.50       Matches:        265
Percent Similarity:       78.87%        Conservative:   0
Best Local Similarity:    78.87%        Mismatches:     4
Query Match:              50.08%        Indels:         69
```

```
DB:                         12             Gaps:          1

US-09-886-954A-1 (1-477) x BI757409 (1-889)

Qy        143 ValTyrIleSerGluIleAlaTyrProAlaValArgGlyLeuLeuGlySerCysValGln 162
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         38 GTCTACATCTCCGAAATCGCCTACCCAGCAGTCCGGGGGTTGCTCGGCTCCTGTGTGCAG 97

Qy        163 LeuMetValValValGlyIleLeuLeuAlaTyrLeuAlaGlyTrpValLeuGluTrpArg 182
              |||||||||||||||||||||||||||||||||||||||
Db         98 CTAATGGTCGTCGTCGGCATCCTCCTGGCCTACCTGGCA--------------------- 136

Qy        183 TrpLeuAlaValLeuGlyCysValProProSerLeuMetLeuLeuLeuMetCysPheMet 202

Db        136 ------------------------------------------------------------ 136

Qy        203 ProGluThrProArgPheLeuLeuThrGlnHisArgArgGlnGluAlaMetAlaAlaLeu 222

Db        136 ------------------------------------------------------------ 136

Qy        223 ArgPheLeuTrpGlySerGluGlnGlyTrpGluAspProProIleGlyAlaGluGlnSer 242

Db        137 ---------------------------------------------------------GAG 139

Qy        243 -PheHisLeuAlaLeuLeuArgGlnProGlyIleTyrLysProPheIleIleGlyValSe 262
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        140 CTTTCACCTGGCCCTGCTGCGGCAGCCCGGCATCTACAAGCCCTTCATCATCGGCGTCTC 199

Qy        262 rLeuMetAlaPheGlnGlnLeuSerGlyValAsnAlaValMetPheTyrAlaGluThrIl 282
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        200 CCTGATGGCCTTCCAGCAGCTGTCGGGGGTCAACGCCGTCATGTTCTATGCAGAGACCAT 259

Qy        282 ePheGluGluAlaLysPheLysAspSerSerLeuAlaSerValValValGlyValIleGl 302
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        260 CTTTGAAGAGGCCAAGTTCAAGGACAGCAGCCTGGCCTCGGTCGTCGTGGGTGTCATCCA 319

Qy        302 nValLeuPheThrAlaValAlaAlaLeuIleMetAspArgAlaGlyArgArgLeuLeuLe 322
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        320 GGTGCTGTTCACAGCTGTGGCGGCTCTCATCATGGACAGAGCAGGGCGGAGGCTGCTCCT 379

Qy        322 uValLeuSerGlyValValMetValPheSerThrSerAlaPheGlyAlaTyrPheLysLe 342
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        380 GGTCTTGTCAGGTGTGGTCATGGTGTTCAGCACGAGTGCCTTCGGCGCCTACTTCAAGCT 439

Qy        342 uThrGlnGlyGlyProGlyAsnSerSerHisValAlaIleSerAlaProValSerAlaGl 362
              ||||   |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        440 GACCCA-GGTGGCCCTGGCAACTCCTCGCACGTGGCCATCTCGGCGCCTGTCTCTGCACA 498

Qy        362 nProValAspAlaSerValGlyLeuAlaTrpLeuAlaValGlySerMetCysLeuPheIl 382
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        499 GCCTGTTGATGCCAGCGTGGGGCTGGCCTGGCTGGCCGTGGGCAGCATGTGCCTCTTCAT 558

Qy        382 eAlaGlyPheAlaValGlyTrpGlyProIleProTrpLeuLeuMetSerGluIlePhePr 402
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        559 CGCCGGCTTTGCGGTGGGCTGGGGGCCCATCCCCTGGCTCCTCATGTCAGAGATCTTCCC 618

Qy        402 oLeuHisValLysGlyValAlaThrGlyIleCysValLeuThrAsnTrpLeuMetAlaPh 422
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        619 TCTGCATGTCAAGGGCGTGGCGACAGGCATCTGCGTCCTCACCAACTGGCTCATGGCCTT 678

Qy        422 eLeuValThrLysGluPheSerSerLeuMetGluValLeuArgProTyrGlyAlaPheTr 442
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        679 TCTCGTGACCAAGGAGTTCAGCAGCCTCATGGAGGTCCTCAGGCCCTATGGAGCCTTCTG 738

Qy        442 pLeuAlaSerAlaPheCysIlePheSerValLeuPheThrLeuPheCysValProGluTh 462
              ||||||||||   |||   |||||||||||||||||||||||||||||||||||||||||
Db        739 GCTTGCCTCCG--TTTCTGATCTTCAGTGTCCTTTTCACTTTGTTCTGTGTCCCTGAAAC 796

Qy        462 rLysGlyLysThrLeuGluGlnIleThrAlaHisPheGluGlyArg 477
              ||||||||||||||||||||||||||||||||||||||||||||||
Db        797 TAAAGGAAAGACTCTGGAACAAATCACAGCCCATTTTGAGGGGCGA 842
```

```
Search completed: September 28, 2004, 04:16:06
Job time : 2866 secs
```

```
                    GenCore version 5.1.6
           Copyright (c) 1993 - 2004  Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on:          September 28, 2004, 00:52:18 ; Search time 484 Seconds
                                          (without alignments)
                                          4186.757 Million cell updates/sec

Title:           US-09-886-954A-1
Perfect score:   2457
Sequence:        1 MTPEDPEETQPLLGPPGGSA..........CVPETKGKTLEQITAHFEGR 477

Scoring table:   BLOSUM62
                 Xgapop 10.0 , Xgapext  0.5
                 Ygapop 10.0 , Ygapext  0.5
                 Fgapop  6.0 , Fgapext  7.0
                 Delop   6.0 , Delext   7.0

Searched:        3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters:      6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US09886954/runat_27092004_170631_13378/app_query.fasta_1.647
-DB=N_Geneseq_29Jan04 -QFMT=fastap -SUFFIX=rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09886954_@CGN_1_1_708_@runat_27092004_170631_13378 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :       N_Geneseq_29Jan04:*
                 1:  geneseqn1980s:*
                 2:  geneseqn1990s:*
                 3:  geneseqn2000s:*
                 4:  geneseqn2001as:*
                 5:  geneseqn2001bs:*
                 6:  geneseqn2002s:*
                 7:  geneseqn2003as:*
                 8:  geneseqn2003bs:*
                 9:  geneseqn2003cs:*
                 10:  geneseqn2004s:*

     Pred. No. is the number of results predicted by chance to have a
     score greater than or equal to the score of the result being printed,
     and is derived by analysis of the total score distribution.

                              SUMMARIES
                 %
Result          Query
  No.   Score   Match Length DB  ID                       Description
-----------------------------------------------------------------------------
    1    2457   100.0    1445  7  ABZ24794                 Abz24794 Human sol
    2    2457   100.0    1856  7  ABZ24792                 Abz24792 Human sol
    3    2457   100.0    2080  4  AAD09552                 Aad09552 Human tra
    4    2448    99.6    1873  7  ABZ24793                 Abz24793 Human sol
    5    2448    99.6    2217  4  AAF55865                 Aaf55865 Human GLU
    6  2162.5    88.0    2087  4  AAF55866                 Aaf55866 Rat GLUTX
    7    2150    87.5    2072  4  AAF55867                 Aaf55867 Murine GL
    8  2123.5    86.4    1954  6  ABX34029                 Abx34029 Human can
```

```
      9  1679.5    68.4    1156  6  ABL90126                 Abl90126 Human pol
     10  1554.5    63.3    1461  4  AAD12574                 Aad12574 Human pro
     11   972.5    39.6    2011  4  AAF55871                 Aaf55871 Rat GLUTX
     12   958.5    39.0    2356  7  ADA84076                 Ada84076 Human SLC
     13   958.5    39.0    2356  7  ABZ24791                 Abz24791 Human sol
     14   958.5    39.0    2390  4  AAK94338                 Aak94338 Human ful
     15   958.5    39.0    2487  7  ABZ24790                 Abz24790 Human sol
     16     949    38.6    1541  4  AAF55870                 Aaf55870 Human GLU
     17   904.5    36.8    2171  4  AAK94284                 Aak94284 Human ful
     18   804.5    32.7    1412  4  ABL13883                 Abl13883 Drosophil
     19   788.5    32.1    1724  4  ABL13389                 Abl13389 Drosophil
     20     788    32.1    2514  4  ABL17067                 Abl17067 Drosophil
     21     734    29.9    1662  4  ABL11183                 Abl11183 Drosophil
     22     705    28.7    1632  4  ABL02211                 Abl02211 Drosophil
c    23   701.5    28.6    3716  4  ABL13882                 Abl13882 Drosophil
     24     687    28.0    9218  4  ABL17066                 Abl17066 Drosophil
     25   686.5    27.9    1527  2  AAZ11731                 Aaz11731 Human tra
c    26     676    27.5    4471  4  ABL13388                 Abl13388 Drosophil
     27   670.5    27.3    1741  3  AAC36319                 Aac36319 Arabidops
     28   666.5    27.1    1614  4  ABL02481                 Abl02481 Drosophil
     29   666.5    27.1    3614  4  ABL02480                 Abl02480 Drosophil
     30   662.5    27.0    1683  3  AAC39099                 Aac39099 Arabidops
     31     661    26.9    4078  4  ABL11182                 Abl11182 Drosophil
     32   653.5    26.6    1755  3  AAC40459                 Aac40459 Arabidops
     33   652.5    26.6    1395  3  AAC45857                 Aac45857 Arabidops
     34     647    26.3    1767  4  ABL17113                 Abl17113 Drosophil
     35     647    26.3    1856  4  ABL02213                 Abl02213 Drosophil
     36     642    26.1    1685  3  AAC45298                 Aac45298 Arabidops
     37     642    26.1    1743  3  AAC45290                 Aac45290 Arabidops
c    38     620    25.2    8597  4  ABL02210                 Abl02210 Drosophil
c    39     606    24.7    4695  4  ABL17112                 Abl17112 Drosophil
c    40     606    24.7    7393  4  ABL02212                 Abl02212 Drosophil
     41   598.5    24.4    1395  7  ACA32312                 Aca32312 Prokaryot
     42   591.5    24.1    1395  7  ACA18864                 Aca18864 Prokaryot
     43   591.5    24.1   12354  4  AAS46267                 Aas46267 DNA encod
     44   591.5    24.1   12354  4  AAS46243                 Aas46243 DNA encod
     45   591.5    24.1   12354  4  AAS46261                 Aas46261 DNA encod

                              ALIGNMENTS

RESULT 1
ABZ24794
ID   ABZ24794 standard; cDNA; 1445 BP.
XX
AC   ABZ24794;
XX
DT   07-APR-2003  (first entry)
XX
DE   Human solute carrier type 2A nucleic acid.
XX
KW   Human; solute carrier type 2A; SLC2A; glucose transporter; p53; cancer;
KW   cytostatic; gene therapy; gene; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             7. .1440
FT                   /*tag=  a
FT                   /product= "Human SCL2A"
XX
PN   WO200298467-A1.
XX
PD   12-DEC-2002.
XX
PF   03-JUN-2002; 2002WO-US017419.
XX
PR   05-JUN-2001; 2001US-0296076P.
PR   10-OCT-2001; 2001US-0328605P.
PR   15-FEB-2002; 2002US-0357253P.
XX
PA   (EXEL-) EXELIXIS INC.
XX
```

```
PI   Friedman L,  Plowman GD,  Belvin M,  Francis-Lang H,  Li D,  Funke RP;
XX
DR   WPI; 2003-201283/19.
DR   P-PSDB; ABP58364.
XX
PT   Identifying candidate p53 pathway modulating agent as therapeutic target
PT   for disorders of defective p53 function e.g. cancer, by assaying purified
PT   solute carrier family 2 (SLC2A) polypeptide or nucleic acid with a test
PT   agent.
XX
PS   Disclosure; Page 45; 58pp; English.
XX
CC   The present sequence is that of a human solute carrier type 2A (SLC2A)
CC   nucleic acid. Genetic screens were designed to identify modifiers of the
CC   p53 pathway in Drosophila in which p53 was overexpressed. Human
CC   orthologues (polynucleotides and polypeptides) of one such modifier were
CC   then identified, including the present nucleic acid. SLC2As are glucose
CC   transporter proteins with sugar transporter domains. SLC2A nucleic acids
CC   and polypeptides are attractive drug targets for the treatment of
CC   pathologies associated with a defective p53 signalling pathway, such as
CC   cancer. The invention provides in vitro and in vivo methods of assessing
CC   SLC2A function. Modulation of an SLC2A or its binding partners is useful
CC   for understanding the association of the p53 pathway and its members in
CC   normal and disease conditions and for developing diagnostic and
CC   therapeutic modalities of p53-related pathologies. SLC2A-modulating
CC   agents that act by inhibiting or enhancing SLC2A expression, directly or
CC   indirectly, e.g. by affecting an SLC2A function such as transport or
CC   binding activity, can be identified using methods provided. Modulators
CC   include small molecules, nucleic acids, antibodies, antisense
CC   oligonucleotides and phosphothioate morpholino oligomers (claimed)
XX
SQ   Sequence 1445 BP; 182 A; 500 C; 456 G; 307 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              1.02e-204     Length:        1445
Score:                  2457.00       Matches:       477
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     7             Gaps:          0

US-09-886-954A-1 (1-477) x ABZ24794 (1-1445)

Qy          1 MetThrProGluAspProGluGluThrGlnProLeuLeuGlyProProGlyGlySerAla 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          7 ATGACGCCCGAGGACCCAGAGGAAACCCAGCCGCTTCTGGGGCCTCCTGGCGGCAGCGCG 66

Qy         21 ProArgGlyArgArgValPheLeuAlaAlaPheAlaAlaAlaLeuGlyProLeuSerPhe 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         67 CCCCGCGGCCGCCGCGTCTTCCTCGCCGCCTTCGCCGCTGCCCTGGGCCCACTCAGCTTC 126

Qy         41 GlyPheAlaLeuGlyTyrSerSerProAlaIleProSerLeuGlnArgAlaAlaProPro 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        127 GGCTTCGCGCTCGGCTACAGCTCCCCGGCCATCCCTAGCCTGCAGCGCGCCGCGCCCCCG 186

Qy         61 AlaProArgLeuAspAspAlaAlaAlaSerTrpPheGlyAlaValValThrLeuGlyAla 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        187 GCCCCGCGCCTGGACGACGCCGCCGCCTCCTGGTTCGGGGCTGTCGTGACCCTGGGTGCC 246

Qy         81 AlaAlaGlyGlyValLeuGlyGlyTrpLeuValAspArgAlaGlyArgLysLeuSerLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        247 GCGGCGGGGGGAGTGCTGGGCGGCTGGCTGGTGGACCGCGCCGGGCGCAAGCTGAGCCTC 306

Qy        101 LeuLeuCysSerValProPheValAlaGlyPheAlaValIleThrAlaAlaGlnAspVal 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        307 TTGCTGTGCTCCGTGCCCTTCGTGGCCGGCTTTGCCGTCATCACCGCGGCCCAGGACGTG 366

Qy        121 TrpMetLeuLeuGlyGlyArgLeuLeuThrGlyLeuAlaCysGlyValAlaSerLeuVal 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        367 TGGATGCTGCTGGGGGGCCGCCTCCTCACCGGCCTGGCCTGCGGTGTTGCCTCCCTAGTG 426

Qy        141 AlaProValTyrIleSerGluIleAlaTyrProAlaValArgGlyLeuLeuGlySerCys 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        427 GCCCCGGTCTACATCTCCGAAATCGCCTACCCAGCAGTCCGGGGGTTGCTCGGCTCCTGT 486

Qy        161 ValGlnLeuMetValValValGlyIleLeuLeuAlaTyrLeuAlaGlyTrpValLeuGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        487 GTGCAGCTAATGGTCGTCGTCGGCATCCTCCTGGCCTACCTGGCAGGCTGGGTGCTGGAG 546

Qy        181 TrpArgTrpLeuAlaValLeuGlyCysValProProSerLeuMetLeuLeuLeuMetCys 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        547 TGGCGCTGGCTGGCTGTGCTGGGCTGCGTGCCCCCCTCCCTCATGCTGCTTCTCATGTGC 606

Qy        201 PheMetProGluThrProArgPheLeuLeuThrGlnHisArgArgGlnGluAlaMetAla 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        607 TTCATGCCCGAGACCCCGCGCTTCCTGCTGACTCAGCACAGGCGCCAGGAGGCCATGGCC 666

Qy        221 AlaLeuArgPheLeuTrpGlySerGluGlnGlyTrpGluAspProProIleGlyAlaGlu 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        667 GCCCTGCGGTTCCTGTGGGGCTCCGAGCAGGGCTGGGAAGACCCCCCCATCGGGGCTGAG 726

Qy        241 GlnSerPheHisLeuAlaLeuLeuArgGlnProGlyIleTyrLysProPheIleIleGly 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        727 CAGAGCTTTCACCTGGCCCTGCTGCGGCAGCCCGGCATCTACAAGCCCTTCATCATCGGC 786

Qy        261 ValSerLeuMetAlaPheGlnGlnLeuSerGlyValAsnAlaValMetPheTyrAlaGlu 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        787 GTCTCCCTGATGGCCTTCCAGCAGCTGTCGGGGGTCAACGCCGTCATGTTCTATGCAGAG 846

Qy        281 ThrIlePheGluGluAlaLysPheLysAspSerSerLeuAlaSerValValValGlyVal 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        847 ACCATCTTTGAAGAGGCCAAGTTCAAGGACAGCAGCCTGGCCTCGGTCGTCGTGGGTGTC 906

Qy        301 IleGlnValLeuPheThrAlaValAlaAlaLeuIleMetAspArgAlaGlyArgArgLeu 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        907 ATCCAGGTGCTGTTCACAGCTGTGGCGGCTCTCATCATGGACAGAGCAGGGCGGAGGCTG 966

Qy        321 LeuLeuValLeuSerGlyValValMetValPheSerThrSerAlaPheGlyAlaTyrPhe 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        967 CTCCTGGTCTTGTCAGGTGTGGTCATGGTGTTCAGCACGAGTGCCTTCGGCGCCTACTTC 1026

Qy        341 LysLeuThrGlnGlyGlyProGlyAsnSerSerHisValAlaIleSerAlaProValSer 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1027 AAGCTGACCCAGGGTGGCCCTGGCAACTCCTCGCACGTGGCCATCTCGGCGCCTGTCTCT 1086

Qy        361 AlaGlnProValAspAlaSerValGlyLeuAlaTrpLeuAlaValGlySerMetCysLeu 380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1087 GCACAGCCTGTTGATGCCAGCGTGGGGCTGGCCTGGCTGGCCGTGGGCAGCATGTGCCTC 1146

Qy        381 PheIleAlaGlyPheAlaValGlyTrpGlyProIleProTrpLeuLeuMetSerGluIle 400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1147 TTCATCGCCGGCTTTGCGGTGGGCTGGGGGCCCATCCCCTGGCTCCTCATGTCAGAGATC 1206

Qy        401 PheProLeuHisValLysGlyValAlaThrGlyIleCysValLeuThrAsnTrpLeuMet 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1207 TTCCCTCTGCATGTCAAGGGCGTGGCGACAGGCATCTGCGTCCTCACCAACTGGCTCATG 1266

Qy        421 AlaPheLeuValThrLysGluPheSerSerLeuMetGluValLeuArgProTyrGlyAla 440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1267 GCCTTTCTCGTGACCAAGGAGTTCAGCAGCCTCATGGAGGTCCTCAGGCCCTATGGAGCC 1326

Qy        441 PheTrpLeuAlaSerAlaPheCysIlePheSerValLeuPheThrLeuPheCysValPro 460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1327 TTCTGGCTTGCCTCCGCTTTCTGCATCTTCAGTGTCCTTTTCACTTTGTTCTGTGTCCCT 1386

Qy        461 GluThrLysGlyLysThrLeuGluGlnIleThrAlaHisPheGluGlyArg 477
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1387 GAAACTAAAGGAAAGACTCTGGAACAAATCACAGCCCATTTTGAGGGGCGA 1437

RESULT 2
ABZ24792
ID   ABZ24792 standard; cDNA; 1856 BP.
XX
AC   ABZ24792;
XX
DT   07-APR-2003  (first entry)
```

```
XX
DE   Human solute carrier type 2A nucleic acid 13642003.
XX
KW   Human; solute carrier type 2A; SLC2A; glucose transporter; p53; cancer;
KW   cytostatic; gene therapy; gene; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             4..1437
FT                   /*tag=  a
FT                   /product= "Human SCL2A"
XX
PN   WO200298467-A1.
XX
PD   12-DEC-2002.
XX
PF   03-JUN-2002; 2002WO-US017419.
XX
PR   05-JUN-2001; 2001US-0296076P.
PR   10-OCT-2001; 2001US-0328605P.
PR   15-FEB-2002; 2002US-0357253P.
XX
PA   (EXEL-) EXELIXIS INC.
XX
PI   Friedman L,  Plowman GD,  Belvin M,  Francis-Lang H,  Li D,  Funke RP;
XX
DR   WPI; 2003-201283/19.
DR   P-PSDB; ABP58364.
XX
PT   Identifying candidate p53 pathway modulating agent as therapeutic target
PT   for disorders of defective p53 function e.g. cancer, by assaying purified
PT   solute carrier family 2 (SLC2A) polypeptide or nucleic acid with a test
PT   agent.
XX
PS   Disclosure; Page 42-43; 58pp; English.
XX
CC   The present sequence is that of human solute carrier type 2A (SLC2A)
CC   nucleic acids 13642003. Genetic screens were designed to identify
CC   modifiers of the p53 pathway in Drosophila in which p53 was
CC   overexpressed. Human orthologues (polynucleotides and polypeptides) of
CC   one such modifier were then identified, including nucleic acid 13642003.
CC   SLC2As are glucose transporter proteins with sugar transporter domains.
CC   SLC2A nucleic acids and polypeptides are attractive drug targets for the
CC   treatment of pathologies associated with a defective p53 signalling
CC   pathway, such as cancer. The invention provides in vitro and in vivo
CC   methods of assessing SLC2A function. Modulation of an SLC2A or its
CC   binding partners is useful for understanding the association of the p53
CC   pathway and its members in normal and disease conditions and for
CC   developing diagnostic and therapeutic modalities of p53-related
CC   pathologies. SLC2A-modulating agents that act by inhibiting or enhancing
CC   SLC2A expression, directly or indirectly, e.g. by affecting an SLC2A
CC   function such as transport or binding activity, can be identified using
CC   methods provided. Modulators include small molecules, nucleic acids,
CC   antibodies, antisense oligonucleotides and phosphothioate morpholino
CC   oligomers (claimed)
XX
SQ   Sequence 1856 BP; 262 A; 624 C; 573 G; 397 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              1.42e-204       Length:         1856
Score:                  2457.00         Matches:        477
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     7               Gaps:           0

US-09-886-954A-1 (1-477) x ABZ24792 (1-1856)

Qy          1 MetThrProGluAspProGluGluThrGlnProLeuLeuGlyProProGlySerAla 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          4 ATGACGCCCGAGGACCCAGAGGAAACCCAGCCGCTTCTGGGGCCTCCTGGCGGCAGCGCG 63

Qy         21 ProArgGlyArgArgValPheLeuAlaAlaPheAlaAlaAlaLeuGlyProLeuSerPhe 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         64 CCCCGCGGCCGCCGCGTCTTCCTCGCCGCCTTCGCCGCTGCCCTGGGCCCACTCAGCTTC 123

Qy         41 GlyPheAlaLeuGlyTyrSerSerProAlaIleProSerLeuGlnArgAlaAlaProPro 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        124 GGCTTCGCGCTCGGCTACAGCTCCCCGGCCATCCCTAGCCTGCAGCGCGCCGCGCCCCCG 183

Qy         61 AlaProArgLeuAspAspAlaAlaAlaSerTrpPheGlyAlaValValThrLeuGlyAla 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        184 GCCCCGCGCCTGGACGACGCCGCCGCCTCCTGGTTCGGGGCTGTCGTGACCCTGGGTGCC 243

Qy         81 AlaAlaGlyGlyValLeuGlyGlyTrpLeuValAspArgAlaGlyArgLysLeuSerLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        244 GCGGCGGGGGGAGTGCTGGGCGGCTGGCTGGTGGACCGCGCCGGGCGCAAGCTGAGCCTC 303

Qy        101 LeuLeuCysSerValProPheValAlaGlyPheAlaValIleThrAlaAlaGlnAspVal 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        304 TTGCTGTGCTCCGTGCCCTTCGTGGCCGGCTTTGCCGTCATCACCGCGGCCCAGGACGTG 363

Qy        121 TrpMetLeuLeuGlyGlyArgLeuLeuThrGlyLeuAlaCysGlyValAlaSerLeuVal 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        364 TGGATGCTGCTGGGGGGCCGCCTCCTCACCGGCCTGGCCTGCGGTGTTGCCTCCCTAGTG 423

Qy        141 AlaProValTyrIleSerGluIleAlaTyrProAlaValArgGlyLeuLeuGlySerCys 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        424 GCCCCGGTCTACATCTCCGAAATCGCCTACCCAGCAGTCCGGGGGTTGCTCGGCTCCTGT 483

Qy        161 ValGlnLeuMetValValValGlyIleLeuLeuAlaTyrLeuAlaGlyTrpValLeuGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        484 GTGCAGCTAATGGTCGTCGTCGGCATCCTCCTGGCCTACCTGGCAGGCTGGGTGCTGGAG 543

Qy        181 TrpArgTrpLeuAlaValLeuGlyCysValProProSerLeuMetLeuLeuLeuMetCys 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        544 TGGCGCTGGCTGGCTGTGCTGGGCTGCGTGCCCCCCTCCCTCATGCTGCTTCTCATGTGC 603

Qy        201 PheMetProGluThrProArgPheLeuLeuThrGlnHisArgArgGlnGluAlaMetAla 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        604 TTCATGCCCGAGACCCCGCGCTTCCTGCTGACTCAGCACAGGCGCCAGGAGGCCATGGCC 663

Qy        221 AlaLeuArgPheLeuTrpGlySerGluGlnGlyTrpGluAspProProIleGlyAlaGlu 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        664 GCCCTGCGGTTCCTGTGGGGCTCCGAGCAGGGCTGGGAAGACCCCCCCATCGGGGCTGAG 723

Qy        241 GlnSerPheHisLeuAlaLeuLeuArgGlnProGlyIleTyrLysProPheIleIleGly 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        724 CAGAGCTTTCACCTGGCCCTGCTGCGGCAGCCCGGCATCTACAAGCCCTTCATCATCGGC 783

Qy        261 ValSerLeuMetAlaPheGlnGlnLeuSerGlyValAsnAlaValMetPheTyrAlaGlu 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        784 GTCTCCCTGATGGCCTTCCAGCAGCTGTCGGGGGTCAACGCCGTCATGTTCTATGCAGAG 843

Qy        281 ThrIlePheGluGluAlaLysPheLysAspSerSerLeuAlaSerValValValGlyVal 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        844 ACCATCTTTGAAGAGGCCAAGTTCAAGGACAGCAGCCTGGCCTCGGTCGTCGTGGGTGTC 903

Qy        301 IleGlnValLeuPheThrAlaValAlaAlaLeuIleMetAspArgAlaGlyArgArgLeu 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        904 ATCCAGGTGCTGTTCACAGCTGTGGCGGCTCTCATCATGGACAGAGCAGGGCGGAGGCTG 963

Qy        321 LeuLeuValLeuSerGlyValValMetValPheSerThrSerAlaPheGlyAlaTyrPhe 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        964 CTCCTGGTCTTGTCAGGTGTGGTCATGGTGTTCAGCACGAGTGCCTTCGGCGCCTACTTC 1023

Qy        341 LysLeuThrGlnGlyGlyProGlyAsnSerSerHisValAlaIleSerAlaProValSer 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1024 AAGCTGACCCAGGGTGGCCCTGGCAACTCCTCGCACGTGGCCATCTCGGCGCCTGTCTCT 1083

Qy        361 AlaGlnProValAspAlaSerValGlyLeuAlaTrpLeuAlaValGlySerMetCysLeu 380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1084 GCACAGCCTGTTGATGCCAGCGTGGGGCTGGCCTGGCTGGCCGTGGGCAGCATGTGCCTC 1143

Qy        381 PheIleAlaGlyPheAlaValGlyTrpGlyProIleProTrpLeuLeuMetSerGluIle 400
```

```
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1144 TTCATCGCCGGCTTTGCGGTGGGCTGGGGGCCCATCCCCTGGCTCCTCATGTCAGAGATC 1203

Qy        401 PheProLeuHisValLysGlyValAlaThrGlyIleCysValLeuThrAsnTrpLeuMet 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1204 TTCCCTCTGCATGTCAAGGGCGTGGCGACAGGCATCTGCGTCCTCACCAACTGGCTCATG 1263

Qy        421 AlaPheLeuValThrLysGluPheSerSerLeuMetGluValLeuArgProTyrGlyAla 440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1264 GCCTTTCTCGTGACCAAGGAGTTCAGCAGCCTCATGGAGGTCCTCAGGCCCTATGGAGCC 1323

Qy        441 PheTrpLeuAlaSerAlaPheCysIlePheSerValLeuPheThrLeuPheCysValPro 460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1324 TTCTGGCTTGCCTCCGCTTTCTGCATCTTCAGTGTCCTTTTCACTTTGTTCTGTGTCCCT 1383

Qy        461 GluThrLysGlyLysThrLeuGluGlnIleThrAlaHisPheGluGlyArg 477
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1384 GAAACTAAAGGAAAGACTCTGGAACAAATCACAGCCCATTTTGAGGGGCGA 1434

RESULT 3
AAD09552
ID   AAD09552 standard; cDNA; 2080 BP.
XX
AC   AAD09552;
XX
DT   10-SEP-2001  (first entry)
XX
DE   Human transporter and ion channel-1 (TRICH-1) cDNA.
XX
KW   Human; transporter and ion channel-1; TRICH-1; vaccine; cystic fibrosis;
KW   gene therapy; amyotrophic lateral sclerosis; amnesia; muscular dystrophy;
KW   hypertension; angina; neurological disorder; asthma; bipolar disorder;
KW   dementia; depression; Alzheimer's disease; epilepsy; mood; arrhythmia;
KW   Pick's disease; ischaemic cerebrovascular disease; AIDS; anxiety; stroke;
KW   Huntington's disease; Parkinson's disease; cerebral neoplasm; allergy;
KW   demyelinating disease; mental disorder; Schizophrenia; polymyositis;
KW   muscle disorder; cardiomyopathy; cataract; myocarditis; Grave's disease;
KW   dermatomyositis; diabetes mellitus; immunological disorder; psoriasis;
KW   rheumatoid arthritis; Sjogren's syndrome; systemic lupus erythematosus;
KW   sickle cell anaemia; Wilson's disease; infertility; Cushing's disease;
KW   scleroderma; pulmonary artery stenosis; nootropic; Addison's disease;
KW   malabsorption syndrome; hypercholesterolaemia; cancer; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             46. .1479
FT                   /*tag=  a
FT                   /product= "Human TRICH-1 protein"
XX
PN   WO200146258-A2.
XX
PD   28-JUN-2001.
XX
PF   22-DEC-2000; 2000WO-US035095.
XX
PR   23-DEC-1999;   99US-0172000P.
PR   14-JAN-2000; 2000US-0176083P.
PR   21-JAN-2000; 2000US-0177332P.
PR   28-JAN-2000; 2000US-0178572P.
PR   02-FEB-2000; 2000US-0179758P.
PR   10-FEB-2000; 2000US-0181625P.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Baughn MR,  Burford N,  Au-Young J,  Lu DAM,  Yang J,  Reddy R;
PI   Lal P,  Hillman JL,  Azimzai Y,  Yue H,  Nguyen DB,  Yao MG,  Gandhi AR;
PI   Tang YT,  Khan FA;
XX
DR   WPI; 2001-418042/44.
DR   P-PSDB; AAE04888.
XX
PT   Novel human transporter and ion channel proteins useful for treating and
PT   preventing transport, neurological, muscle and immunological disorders.
XX
PS   Claim 5; Page 140-141; 160pp; English.
XX
CC   The present sequence is transporter and ion channel-1 (TRICH-1) cDNA.
CC   TRICH is used as vaccine. TRICH is useful for treating a disease or
CC   condition associated with decreased expression of functional TRICH, such
CC   as transport disorder including amyotrophic lateral sclerosis, cystic
CC   fibrosis, Becker's muscular dystrophy, Charcot-Marie Tooth disease,
CC   Duchenne muscular dystrophy, angina and hypertension, neurological
CC   disorders including Alzheimer's disease, amnesia, bipolar disorder,
CC   dementia, depression, epilepsy, ischaemic cerebrovascular disease,
CC   stroke, cerebral neoplasms, Pick's disease, Huntington's disease and
CC   Parkinson's disease, demyelinating diseases, mental disorders including
CC   mood, anxiety, Schizophrenia and seasonal affective disorder, muscle
CC   disorder including cardiomyopathy, myocarditis, polymyositis,
CC   dermatomyositis, arrhythmias and asthma and immunological disorders
CC   including AIDS, adult respiratory distress syndrome (ARDS), allergies,
CC   anaemia, diabetes mellitus, rheumatoid arthritis, scleroderma, Sjogren's
CC   syndrome, systemic lupus erythematosus and other diseases including
CC   sickle cell anaemia, Wilson's disease, cataracts, infertility, pulmonary
CC   artery stenosis, Grave's disease, Cushing's disease, Addison's disease,
CC   glucose-galactose malabsorption syndrome, hypercholesterolaemia, cancers
CC   psoriasis and viral, bacterial, fungal, helminthic and protozoal
CC   infections. TRICH DNA is useful in gene therapy and in diagnostic
CC   purposes
XX
SQ   Sequence 2080 BP; 300 A; 688 C; 632 G; 460 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              1.65e-204      Length:         2080
Score:                  2457.00        Matches:        477
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     4              Gaps:           0

US-09-886-954A-1 (1-477) x AAD09552 (1-2080)

Qy          1 MetThrProGluAspProGluGluThrGlnProLeuLeuGlyProProGlyGlySerAla 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         46 ATGACGCCCGAGGACCCAGAGGAAACCCAGCCGCTTCTGGGGCCTCCTGGCGGCAGCGCG 105

Qy         21 ProArgGlyArgArgValPheLeuAlaAlaPheAlaAlaAlaLeuGlyProLeuSerPhe 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        106 CCCCGCGGCCGCCGCGTCTTCCTCGCCGCCTTCGCCGCTGCCCTGGGCCCACTCAGCTTC 165

Qy         41 GlyPheAlaLeuGlyTyrSerSerProAlaIleProSerLeuGlnArgAlaAlaProPro 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        166 GGCTTCGCGCTCGGCTACAGCTCCCCGGCCATCCCTAGCCTGCAGCGCGCCGCGCCCCCG 225

Qy         61 AlaProArgLeuAspAspAlaAlaAlaSerTrpPheGlyAlaValValThrLeuGlyAla 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        226 GCCCCGCGCCTGGACGACGCCGCCGCCTCCTGGTTCGGGGCTGTCGTGACCCTGGGTGCC 285

Qy         81 AlaAlaGlyGlyValLeuGlyGlyTrpLeuValAspArgAlaGlyArgLysLeuSerLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        286 GCGGCGGGGGGAGTGCTGGGCGGCTGGCTGGTGGACCGCGCCGGGCGCAAGCTGAGCCTC 345

Qy        101 LeuLeuCysSerValProPheValAlaGlyPheAlaValIleThrAlaAlaGlnAspVal 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        346 TTGCTGTGCTCCGTGCCCTTCGTGGCCGGCTTTGCCGTCATCACCGCGGCCCAGGACGTG 405

Qy        121 TrpMetLeuLeuGlyGlyArgLeuLeuThrGlyLeuAlaCysGlyValAlaSerLeuVal 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        406 TGGATGCTGCTGGGGGGCCGCCTCCTCACCGGCCTGGCCTGCGGTGTTGCCTCCCTAGTG 465

Qy        141 AlaProValTyrIleSerGluIleAlaTyrProAlaValArgGlyLeuLeuGlySerCys 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        466 GCCCCGGTCTACATCTCCGAAATCGCCTACCCAGCAGTCCGGGGGTTGCTCGGCTCCTGT 525

Qy        161 ValGlnLeuMetValValValGlyIleLeuLeuAlaTyrLeuAlaGlyTrpValLeuGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        526 GTGCAGCTAATGGTCGTCGTCGGCATCCTCCTGGCCTACCTGGCAGGCTGGGTGCTGGAG 585

Qy        181 TrpArgTrpLeuAlaValLeuGlyCysValProProSerLeuMetLeuLeuLeuMetCys 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        586 TGGCGCTGGCTGGCTGTGCTGGGCTGCGTGCCCCCCTCCCTCATGCTGCTTCTCATGTGC 645

Qy        201 PheMetProGluThrProArgPheLeuLeuThrGlnHisArgArgGlnGluAlaMetAla 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        646 TTCATGCCCGAGACCCCGCGCTTCCTGCTGACTCAGCACAGGCGCCAGGAGGCCATGGCC 705

Qy        221 AlaLeuArgPheLeuTrpGlySerGluGlnGlyTrpGluAspProProIleGlyAlaGlu 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        706 GCCCTGCGGTTCCTGTGGGGCTCCGAGCAGGGCTGGGAAGACCCCCCCATCGGGGCTGAG 765

Qy        241 GlnSerPheHisLeuAlaLeuLeuArgGlnProGlyIleTyrLysProPheIleIleGly 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        766 CAGAGCTTTCACCTGGCCCTGCTGCGGCAGCCCGGCATCTACAAGCCCTTCATCATCGGC 825

Qy        261 ValSerLeuMetAlaPheGlnGlnLeuSerGlyValAsnAlaValMetPheTyrAlaGlu 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        826 GTCTCCCTGATGGCCTTCCAGCAGCTGTCGGGGGTCAACGCCGTCATGTTCTATGCAGAG 885

Qy        281 ThrIlePheGluGluAlaLysPheLysAspSerSerLeuAlaSerValValValGlyVal 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        886 ACCATCTTTGAAGAGGCCAAGTTCAAGGACAGCAGCCTGGCCTCGGTCGTCGTGGGTGTC 945

Qy        301 IleGlnValLeuPheThrAlaValAlaAlaLeuIleMetAspArgAlaGlyArgArgLeu 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        946 ATCCAGGTGCTGTTCACAGCTGTGGCGGCTCTCATCATGGACAGAGCAGGGCGGAGGCTG 1005

Qy        321 LeuLeuValLeuSerGlyValValMetValPheSerThrSerAlaPheGlyAlaTyrPhe 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1006 CTCCTGGTCTTGTCAGGTGTGGTCATGGTGTTCAGCACGAGTGCCTTCGGCGCCTACTTC 1065

Qy        341 LysLeuThrGlnGlyGlyProGlyAsnSerSerHisValAlaIleSerAlaProValSer 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1066 AAGCTGACCCAGGGTGGCCCTGGCAACTCCTCGCACGTGGCCATCTCGGCGCCTGTCTCT 1125

Qy        361 AlaGlnProValAspAlaSerValGlyLeuAlaTrpLeuAlaValGlySerMetCysLeu 380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1126 GCACAGCCTGTTGATGCCAGCGTGGGGCTGGCCTGGCTGGCCGTGGGCAGCATGTGCCTC 1185

Qy        381 PheIleAlaGlyPheAlaValGlyTrpGlyProIleProTrpLeuLeuMetSerGluIle 400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1186 TTCATCGCCGGCTTTGCGGTGGGCTGGGGGCCCATCCCCTGGCTCCTCATGTCAGAGATC 1245

Qy        401 PheProLeuHisValLysGlyValAlaThrGlyIleCysValLeuThrAsnTrpLeuMet 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1246 TTCCCTCTGCATGTCAAGGGCGTGGCGACAGGCATCTGCGTCCTCACCAACTGGCTCATG 1305

Qy        421 AlaPheLeuValThrLysGluPheSerSerLeuMetGluValLeuArgProTyrGlyAla 440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1306 GCCTTTCTCGTGACCAAGGAGTTCAGCAGCCTCATGGAGGTCCTCAGGCCCTATGGAGCC 1365

Qy        441 PheTrpLeuAlaSerAlaPheCysIlePheSerValLeuPheThrLeuPheCysValPro 460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1366 TTCTGGCTTGCCTCCGCTTTCTGCATCTTCAGTGTCCTTTTCACTTTGTTCTGTGTCCCT 1425

Qy        461 GluThrLysGlyLysThrLeuGluGlnIleThrAlaHisPheGluGlyArg 477
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1426 GAAACTAAAGGAAAGACTCTGGAACAAATCACAGCCCATTTTGAGGGGCGA 1476

RESULT 4
ABZ24793
ID   ABZ24793 standard; cDNA; 1873 BP.
XX
AC   ABZ24793;
XX
DT   07-APR-2003  (first entry)
XX
DE   Human solute carrier type 2A nucleic acid 7657680.
XX
KW   Human; solute carrier type 2A; SLC2A; glucose transporter; p53; cancer;
KW   cytostatic; gene therapy; gene; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             4. .1437
FT                   /*tag=  a
FT                   /product= "Human SCL2A"
XX
PN   WO200298467-A1.
XX
PD   12-DEC-2002.
XX
PF   03-JUN-2002; 2002WO-US017419.
XX
PR   05-JUN-2001; 2001US-0296076P.
PR   10-OCT-2001; 2001US-0328605P.
PR   15-FEB-2002; 2002US-0357253P.
XX
PA   (EXEL-) EXELIXIS INC.
XX
PI   Friedman L,  Plowman GD,  Belvin M,  Francis-Lang H,  Li D,  Funke RP;
XX
DR   WPI; 2003-201283/19.
DR   P-PSDB; ABP58364.
XX
PT   Identifying candidate p53 pathway modulating agent as therapeutic target
PT   for disorders of defective p53 function e.g. cancer, by assaying purified
PT   solute carrier family 2 (SLC2A) polypeptide or nucleic acid with a test
PT   agent.
XX
PS   Disclosure; Page 43-44; 58pp; English.
XX
CC   The present sequence is that of human solute carrier type 2A (SLC2A)
CC   nucleic acids 7657680. Genetic screens were designed to identify
CC   modifiers of the p53 pathway in Drosophila in which p53 was
CC   overexpressed. Human orthologues (polynucleotides and polypeptides) of
CC   one such modifier were then identified, including nucleic acid 7657680.
CC   SLC2As are glucose transporter proteins with sugar transporter domains.
CC   SLC2A nucleic acids and polypeptides are attractive drug targets for the
CC   treatment of pathologies associated with a defective p53 signalling
CC   pathway, such as cancer. The invention provides in vitro and in vivo
CC   methods of assessing SLC2A function. Modulation of an SLC2A or its
CC   binding partners is useful for understanding the association of the p53
CC   pathway and its members in normal and disease conditions and for
CC   developing diagnostic and therapeutic modalities of p53-related
CC   pathologies. SLC2A-modulating agents that act by inhibiting or enhancing
CC   SLC2A expression, directly or indirectly, e.g. by affecting an SLC2A
CC   function such as transport or binding activity, can be identified using
CC   methods provided. Modulators include small molecules, nucleic acids,
CC   antibodies, antisense oligonucleotides and phosphothioate morpholino
CC   oligomers (claimed)
XX
SQ   Sequence 1873 BP; 279 A; 621 C; 573 G; 400 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:               8.77e-204     Length:        1873
Score:                   2448.00       Matches:       475
Percent Similarity:      99.79%        Conservative:  1
Best Local Similarity:   99.58%        Mismatches:    1
Query Match:             99.63%        Indels:        0
DB:                      7             Gaps:          0

US-09-886-954A-1 (1-477) x ABZ24793 (1-1873)

Qy          1 MetThrProGluAspProGluGluThrGlnProLeuLeuGlyProProGlyGlySerAla 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          4 ATGACGCCCGAGGACCCAGAGGAAACCCAGCCGCTTCTGGGGCCTCCTGGCGGCAGCGCG 63

Qy         21 ProArgGlyArgArgValPheLeuAlaAlaPheAlaAlaAlaLeuGlyProLeuSerPhe 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         64 CCCCGCGGCCGCCGCGTCTTCCTCGCCGCCTTCGCCGCTGCCCTGGGCCCACTCAGCTTC 123
```

```
Qy          41 GlyPheAlaLeuGlyTyrSerSerProAlaIleProSerLeuGlnArgAlaAlaProPro 60
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         124 GGCTTCGCGCTCGGCTACAGCTCCCCGGCCATCCCTAGCCTGCAGCGCGCCGCGCCCCCG 183

Qy          61 AlaProArgLeuAspAspAlaAlaAlaSerTrpPheGlyAlaValValThrLeuGlyAla 80
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         184 GCCCCGCGCCTGGACGACGCCGCCGCCTCCTGGTTCGGGGCTGTCGTGACCCTGGGTGCC 243

Qy          81 AlaAlaGlyGlyValLeuGlyGlyTrpLeuValAspArgAlaGlyArgLysLeuSerLeu 100
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         244 GCGGCGGGGGGAGTGCTGGGCGGCTGGCTGGTGGACCGCGCCGGGCGCAAGCTGAGCCTC 303

Qy         101 LeuLeuCysSerValProPheValAlaGlyPheAlaValIleThrAlaAlaGlnAspVal 120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         304 TTGCTGTGCTCCGTGCCCTTCGTGGCCGGCTTTGCCGTCATCACCGCGGCCCAGGACGTG 363

Qy         121 TrpMetLeuLeuGlyGlyArgLeuLeuThrGlyLeuAlaCysGlyValAlaSerLeuVal 140
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         364 TGGATGCTGCTGGGGGGCCGCCTCCTCACCGGCCTGGCCTGCGGTGTTGCCTCCCTAGTG 423

Qy         141 AlaProValTyrIleSerGluIleAlaTyrProAlaValArgGlyLeuLeuGlySerCys 160
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         424 GCCCCGGTCTACATCTCCGAAATCGCCTACCCAGCAGTCCGGGGGTTGCTCGGCTCCTGT 483

Qy         161 ValGlnLeuMetValValValGlyIleLeuLeuAlaTyrLeuAlaGlyTrpValLeuGlu 180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         484 GTGCAGCTAATGGTCGTCGTCGGCATCCTCCTGGCCTACCTGGCAGGCTGGGTGCTGGAG 543

Qy         181 TrpArgTrpLeuAlaValLeuGlyCysValProProSerLeuMetLeuLeuLeuMetCys 200
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         544 TGGCGCTGGCTGGCTGTGCTGGGCTGCGTGCCCCCCTCCCTCATGCTGCTTCTCATGTGC 603

Qy         201 PheMetProGluThrProArgPheLeuLeuThrGlnHisArgArgGlnGluAlaMetAla 220
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         604 TTCATGCCCGAGACCCCGCGCTTCCTGCTGACTCAGCACAGGCGCCAGGAGGCCATGGCC 663

Qy         221 AlaLeuArgPheLeuTrpGlySerGluGlnGlyTrpGluAspProProIleGlyAlaGlu 240
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         664 GCCCTGCGGTTCCTGTGGGGCTCCGAGCAGGGCTGGGAAGACCCCCCCATCGGGGCTGAG 723

Qy         241 GlnSerPheHisLeuAlaLeuLeuArgGlnProGlyIleTyrLysProPheIleIleGly 260
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         724 CAGAGCTTTCACCTGGCCCTGCTGCGGCAGCCCGGCATCTACAAGCCCTTCATCATCGGC 783

Qy         261 ValSerLeuMetAlaPheGlnGlnLeuSerGlyValAsnAlaValMetPheTyrAlaGlu 280
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         784 GTCTCCCTGATGGCCTTCCAGCAGCTGTCGGGGGTCAACGCCGTCATGTTCTATGCAGAG 843

Qy         281 ThrIlePheGluGluAlaLysPheLysAspSerSerLeuAlaSerValValValGlyVal 300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         844 ACCATCTTTGAAGAGGCCAAGTTCAAGGACAGCAGCCTGGCCTCGGTCGTCGTGGGTGTC 903

Qy         301 IleGlnValLeuPheThrAlaValAlaAlaLeuIleMetAspArgAlaGlyArgArgLeu 320
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         904 ATCCAGGTGCTGTTCACAGCTGTGGCGGCTCTCATCATGGACAGAGCAGGGCGGAGGCTG 963

Qy         321 LeuLeuValLeuSerGlyValValMetValPheSerThrSerAlaPheGlyAlaTyrPhe 340
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         964 CTCCTGGTCTTGTCAGGTGTGGTCATGGTGTTCAGCACGAGTGCCTTCGGCGCCTACTTC 1023

Qy         341 LysLeuThrGlnGlyGlyProGlyAsnSerSerHisValAlaIleSerAlaProValSer 360
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1024 AAGCTGACCCAGGGTGGCCCTGGCAACTCCTCGCACGTGGCCATCTCGGCGCCTGTCTCT 1083

Qy         361 AlaGlnProValAspAlaSerValGlyLeuAlaTrpLeuAlaValGlySerMetCysLeu 380
               ||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db        1084 GCACAGCCTGTTGATGCCAGCGTGGGGCTGGCCTGGCTGGCCGTGGGCAACATGTGCCTC 1143

Qy         381 PheIleAlaGlyPheAlaValGlyTrpGlyProIleProTrpLeuLeuMetSerGluIle 400
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1144 TTCATCGCCGGCTTTGCGGTGGGCTGGGGGCCCATCCCCTGGCTCCTCATGTCAGAGATC 1203

Qy         401 PheProLeuHisValLysGlyValAlaThrGlyIleCysValLeuThrAsnTrpLeuMet 420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1204 TTCCCTCTGCATGTCAAGGGCGTGGCGACAGGCATCTGCGTCCTCACCAACTGGCTCATG 1263

Qy         421 AlaPheLeuValThrLysGluPheSerSerLeuMetGluValLeuArgProTyrGlyAla 440
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1264 GCCTTTCTCGTGACCAAGGAGTTCAGCAGCCTCATGGAGGTCCTCAGGCCCTATGGAGCC 1323

Qy         441 PheTrpLeuAlaSerAlaPheCysIlePheSerValLeuPheThrLeuPheCysValPro 460
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1324 TTCTGGCTTGCCTCCGCTTTCTGCATCTTCAGTGTCCTTTTCACTTTGTTCTGTGTCCCT 1383

Qy         461 GluThrLysGlyLysThrLeuGluGlnIleThrAlaHisPheGluGlyArg 477
               |||   |||||||||||||||||||||||||||||||||||||||||||||
Db        1384 GAAATTAAAGGAAAGACTCTGGAACAAATCACAGCCCATTTTGAGGGGCGA 1434

RESULT 5
AAF55865
ID   AAF55865 standard; cDNA; 2217 BP.
XX
AC   AAF55865;
XX
DT   17-APR-2001  (first entry)
XX
DE   Human GLUTX1 coding sequence.
XX
KW   Human; GLUTX; gene therapy; vaccine; hexose transport modulator;
KW   hexose transport disorder; ischaemia; diabetes; hyperglycaemia; ss;
KW   hypoglycaemia; glucose metabolism disorder; neurodegenerative disease.
XX
OS   Homo sapiens.
XX
PN   WO200104145-A2.
XX
PD   18-JAN-2001.
XX
PF   14-JUL-2000; 2000WO-IB001042.
XX
PR   14-JUL-1999;   99US-0143907P.
PR   27-AUG-1999;   99US-0151140P.
PR   23-FEB-2000; 2000US-0184285P.
PR   13-JUL-2000; 2000US-00616132.
XX
PA   (UYLA-) UNIV LAUSANNE.
XX
PI   Thorens B,  Ibberson M,  Uldry M;
XX
DR   WPI; 2001-112615/12.
DR   P-PSDB; AAB66932.
XX
PT   Nucleic acids encoding GLUTX glucose transporter proteins, useful in the
PT   prevention, diagnosis and treatment of hexose transport disorders, e.g.
PT   ischemia and diabetes.
XX
PS   Claim 3; Page 70-71; 124pp; English.
XX
CC   The present invention relates to GLUTX proteins (AAF55865-AAF55871 and
CC   AAB66932-AAB66941). The GLUTX proteins are related to the facultative
CC   glucose carriers GLUT1-GLUT5 and have hexose binding and/or transport
CC   function. The GLUTX proteins may be used in the diagnosis, prevention and
CC   treatment of hexose transport disorders such as ischaemia, diabetes,
CC   hyperglycaemia, hypoglycaemia, a glucose metabolism disorder and/or a
CC   neurodegenerative disease. The present sequence is the coding sequence
CC   for human GLUTX1
XX
SQ   Sequence 2217 BP; 320 A; 750 C; 672 G; 475 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              1.1e-203      Length:        2217
Score:                  2448.00       Matches:       475
Percent Similarity:     99.79%        Conservative:  1
Best Local Similarity:  99.58%        Mismatches:    1
Query Match:            99.63%        Indels:        0
DB:                     4             Gaps:          0
```

```
US-09-886-954A-1 (1-477) x AAF55865 (1-2217)

Qy          1 MetThrProGluAspProGluGluThrGlnProLeuLeuGlyProProGlyGlySerAla 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        348 ATGACGCCCGAGGACCCAGAGGAAACCCAGCCGCTTCTGGGGCCTCCTGGCGGCAGCGCG 407

Qy         21 ProArgGlyArgArgValPheLeuAlaAlaPheAlaAlaAlaLeuGlyProLeuSerPhe 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        408 CCCCGCGGCCGCCGCGTCTTCCTCGCCGCCTTCGCCGCTGCCCTGGGCCCACTCAGCTTC 467

Qy         41 GlyPheAlaLeuGlyTyrSerSerProAlaIleProSerLeuGlnArgAlaAlaProPro 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        468 GGCTTCGCGCTCGGCTACAGCTCCCCGGCCATCCCTAGCCTGCAGCGCGCCGCGCCCCCG 527

Qy         61 AlaProArgLeuAspAspAlaAlaAlaSerTrpPheGlyAlaValValThrLeuGlyAla 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        528 GCCCCGCGCCTGGACGACGCCGCCGCCTCCTGGTTCGGGGCTGTCGTGACCCTGGGTGCC 587

Qy         81 AlaAlaGlyGlyValLeuGlyGlyTrpLeuValAspArgAlaGlyArgLysLeuSerLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        588 GCGGCGGGGGGAGTGCTGGGCGGCTGGCTGGTGGACCGCGCCGGGCGCAAGCTGAGCCTC 647

Qy        101 LeuLeuCysSerValProPheValAlaGlyPheAlaValIleThrAlaAlaGlnAspVal 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        648 TTGCTGTGCTCCGTGCCCTTCGTGGCCGGCTTTGCCGTCATCACCGCGGCCCAGGACGTG 707

Qy        121 TrpMetLeuLeuGlyGlyArgLeuLeuThrGlyLeuAlaCysGlyValAlaSerLeuVal 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        708 TGGATGCTGCTGGGGGGCCGCCTCCTCACCGGCCTGGCCTGCGGTGTTGCCTCCCTAGTG 767

Qy        141 AlaProValTyrIleSerGluIleAlaTyrProAlaValArgGlyLeuLeuGlySerCys 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        768 GCCCCGGTCTACATCTCCGAAATCGCCTACCCAGCAGTCCGGGGGTTGCTCGGCTCCTGT 827

Qy        161 ValGlnLeuMetValValValGlyIleLeuLeuAlaTyrLeuAlaGlyTrpValLeuGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        828 GTGCAGCTAATGGTCGTCGTCGGCATCCTCCTGGCCTACCTGGCAGGCTGGGTGCTGGAG 887

Qy        181 TrpArgTrpLeuAlaValLeuGlyCysValProProSerLeuMetLeuLeuLeuMetCys 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        888 TGGCGCTGGCTGGCTGTGCTGGGCTGCGTGCCCCCCTCCCTCATGCTGCTTCTCATGTGC 947

Qy        201 PheMetProGluThrProArgPheLeuLeuThrGlnHisArgArgGlnGluAlaMetAla 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        948 TTCATGCCCGAGACCCCGCGCTTCCTGCTGACTCAGCACAGGCGCCAGGAGGCCATGGCC 1007

Qy        221 AlaLeuArgPheLeuTrpGlySerGluGlnGlyTrpGluAspProProIleGlyAlaGlu 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1008 GCCCTGCGGTTCCTGTGGGGCTCCGAGCAGGGCTGGGAAGACCCCCCCATCGGGGCTGAG 1067

Qy        241 GlnSerPheHisLeuAlaLeuLeuArgGlnProGlyIleTyrLysProPheIleIleGly 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1068 CAGAGCTTTCACCTGGCCCTGCTGCGGCAGCCCGGCATCTACAAGCCCTTCATCATCGGC 1127

Qy        261 ValSerLeuMetAlaPheGlnGlnLeuSerGlyValAsnAlaValMetPheTyrAlaGlu 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1128 GTCTCCCTGATGGCCTTCCAGCAGCTGTCGGGGGTCAACGCCGTCATGTTCTATGCAGAG 1187

Qy        281 ThrIlePheGluGluAlaLysPheLysAspSerSerLeuAlaSerValValValGlyVal 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1188 ACCATCTTTGAAGAGGCCAAGTTCAAGGACAGCAGCCTGGCCTCGGTCGTCGTGGGTGTC 1247

Qy        301 IleGlnValLeuPheThrAlaValAlaAlaLeuIleMetAspArgAlaGlyArgArgLeu 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1248 ATCCAGGTGCTGTTCACAGCTGTGGCGGCTCTCATCATGGACAGAGCAGGGCGGAGGCTG 1307

Qy        321 LeuLeuValLeuSerGlyValValMetValPheSerThrSerAlaPheGlyAlaTyrPhe 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1308 CTCCTGGTCTTGTCAGGTGTGGTCATGGTGTTCAGCACGAGTGCCTTCGGCGCCTACTTC 1367

Qy        341 LysLeuThrGlnGlyGlyProGlyAsnSerSerHisValAlaIleSerAlaProValSer 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1368 AAGCTGACCCAGGGTGGCCCTGGCAACTCCTCGCACGTGGCCATCTCGGCGCCTGTCTCT 1427

Qy        361 AlaGlnProValAspAlaSerValGlyLeuAlaTrpLeuAlaValGlySerMetCysLeu 380
              ||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db       1428 GCACAGCCTGTTGATGCCAGCGTGGGGCTGGCCTGGCTGGCCGTGGGCAACATGTGCCTC 1487

Qy        381 PheIleAlaGlyPheAlaValGlyTrpGlyProIleProTrpLeuLeuMetSerGluIle 400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1488 TTCATCGCCGGCTTTGCGGTGGGCTGGGGGCCCATCCCCTGGCTCCTCATGTCAGAGATC 1547

Qy        401 PheProLeuHisValLysGlyValAlaThrGlyIleCysValLeuThrAsnTrpLeuMet 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1548 TTCCCTCTGCATGTCAAGGGCGTGGCGACAGGCATCTGCGTCCTCACCAACTGGCTCATG 1607

Qy        421 AlaPheLeuValThrLysGluPheSerSerLeuMetGluValLeuArgProTyrGlyAla 440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1608 GCCTTTCTCGTGACCAAGGAGTTCAGCAGCCTCATGGAGGTCCTCAGGCCCTATGGAGCC 1667

Qy        441 PheTrpLeuAlaSerAlaPheCysIlePheSerValLeuPheThrLeuPheCysValPro 460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1668 TTCTGGCTTGCCTCCGCTTTCTGCATCTTCAGTGTCCTTTTCACTTTGTTCTGTGTCCCT 1727

Qy        461 GluThrLysGlyLysThrLeuGluGlnIleThrAlaHisPheGluGlyArg 477
              |||   |||||||||||||||||||||||||||||||||||||||||||||
Db       1728 GAAATTAAAGGAAAGACTCTGGAACAAATCACAGCCCATTTTGAGGGGCGA 1778
```

```
RESULT 6
AAF55866
ID   AAF55866 standard; cDNA; 2087 BP.
XX
AC   AAF55866;
XX
DT   17-APR-2001  (first entry)
XX
DE   Rat GLUTX1 coding sequence.
XX
KW   Rat; GLUTX; gene therapy; vaccine; hexose transport modulator;
KW   hexose transport disorder; ischaemia; diabetes; hyperglycaemia; ss;
KW   hypoglycaemia; glucose metabolism disorder; neurodegenerative disease.
XX
OS   Rattus sp.
XX
PN   WO200104145-A2.
XX
PD   18-JAN-2001.
XX
PF   14-JUL-2000; 2000WO-IB001042.
XX
PR   14-JUL-1999;   99US-0143907P.
PR   27-AUG-1999;   99US-0151140P.
PR   23-FEB-2000; 2000US-0184285P.
PR   13-JUL-2000; 2000US-00616132.
XX
PA   (UYLA-) UNIV LAUSANNE.
XX
PI   Thorens B,  Ibberson M,  Uldry M;
XX
DR   WPI; 2001-112615/12.
DR   P-PSDB; AAB66933.
XX
PT   Nucleic acids encoding GLUTX glucose transporter proteins, useful in the
PT   prevention, diagnosis and treatment of hexose transport disorders, e.g.
PT   ischemia and diabetes.
XX
PS   Claim 3; Page 71-73; 124pp; English.
XX
CC   The present invention relates to GLUTX proteins (AAF55865-AAF55871 and
CC   AAB66932-AAB66941). The GLUTX proteins are related to the facultative
CC   glucose carriers GLUT1-GLUT5 and have hexose binding and/or transport
CC   function. The GLUTX proteins may be used in the diagnosis, prevention and
CC   treatment of hexose transport disorders such as ischaemia, diabetes,
CC   hyperglycaemia, hypoglycaemia, a glucose metabolism disorder and/or a
CC   neurodegenerative disease. The present sequence is the coding sequence
```

```
CC   for rat GLUTX1
XX
SQ   Sequence 2087 BP; 355 A; 673 C; 580 G; 479 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              8.37e-179    Length:        2087
Score:                  2162.50      Matches:       408
Percent Similarity:     92.26%       Conservative:  33
Best Local Similarity:  85.36%       Mismatches:    36
Query Match:            88.01%       Indels:        1
DB:                     4            Gaps:          1

US-09-886-954A-1 (1-477) x AAF55866 (1-2087)

Qy          1 MetThrProGluAspProGluGluThrGlnProLeuLeuGlyProProGlyGlySerAla 20
              |||:::||||||||||||:::||||||||||||||||||      ||||||      |||
Db         30 ATGTCGCCCGAGGACCCCCAGGAAACACAGCCGCTGTTGCGGTCGCCGGGCGCCAGGGCT 89

Qy         21 ProArgGlyArgArgValPheLeuAlaAlaPheAlaAlaAlaLeuGlyProLeuSerPhe 40
              |||   |||||||||||||||||||||   ||||||||||||||||||||||||||||||
Db         90 CCCGGCGGCCGCCGGGTCTTCCTTGCCACCTTCGCCGCCGCCCTGGGCCCGCTCAGCTTC 149

Qy         41 GlyPheAlaLeuGlyTyrSerSerProAlaIleProSerLeuGlnArgAlaAlaProPro 60
              ||||||||||||||||||||||||||||||||||||||||||:::|||   |||||||||
Db        150 GGCTTCGCGCTTGGCTACAGCTCCCCCGCCATCCCCAGCCTGCGGCGCACCGCACCCCCT 209

Qy         61 AlaProArgLeuAspAspAlaAlaAlaSerTrpPheGlyAlaValValThrLeuGlyAla 80
              |||   ||||||   |||   |||||||||||||||||||||||||||||||||||||||
Db        210 GCCCTACGCCTCGGAGACACTGCGGCCTCCTGGTTCGGGGCCGTCGTGACCCTGGGCGCT 269

Qy         81 AlaAlaGlyGlyValLeuGlyGlyTrpLeuValAspArgAlaGlyArgLysLeuSerLeu 100
              ||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||
Db        270 GCGGCAGGGGGCGTGCTGGGCGGCTGGCTCCTGGACCGTGCAGGGCGCAAGCTGAGCCTC 329

Qy        101 LeuLeuCysSerValProPheValAlaGlyPheAlaValIleThrAlaAlaGlnAspVal 120
              |||||||||:::||||||||||||   ||||||||||||||||||||||||:::||||||
Db        330 CTGCTCTGCACCGTGCCCTTCGTGACCGGTTTTGCTGTCATCACCGCGGCCCGGGATGTG 389

Qy        121 TrpMetLeuLeuGlyGlyArgLeuLeuThrGlyLeuAlaCysGlyValAlaSerLeuVal 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        390 TGGATGCTGCTCGGAGGCCGCCTCCTCACCGGCCTAGCCTGCGGAGTCGCCTCATTAGTG 449

Qy        141 AlaProValTyrIleSerGluIleAlaTyrProAlaValArgGlyLeuLeuGlySerCys 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        450 GCACCGGTCTATATCTCGGAAATCGCCTACCCAGCCGTTCGAGGACTGCTCGGCTCCTGT 509

Qy        161 ValGlnLeuMetValValValGlyIleLeuLeuAlaTyrLeuAlaGlyTrpValLeuGlu 180
              ||||||||||||||||||   ||||||||||||||||||:::||||||||||||||||||
Db        510 GTGCAGCTGATGGTTGTCACTGGCATCCTCTTGGCCTACGTGGCAGGCTGGGTCCTAGAA 569

Qy        181 TrpArgTrpLeuAlaValLeuGlyCysValProProSerLeuMetLeuLeuLeuMetCys 200
              ||||||||||||||||||||||||||||||||||||:::|||||||||||||||||||||
Db        570 TGGCGCTGGTTGGCTGTGCTGGGCTGTGTGCCCCCCACCCTCATGCTGCTGCTCATGTGC 629

Qy        201 PheMetProGluThrProArgPheLeuLeuThrGlnHisArgArgGlnGluAlaMetAla 220
              :::||||||||||||||||||||||||||||||||||||:::   |||||||||||||||
Db        630 TACATGCCCGAGACCCCACGCTTCCTCCTGACGCAACACCAGTACCAGGAGGCCATGGCT 689

Qy        221 AlaLeuArgPheLeuTrpGlySerGluGlnGlyTrpGluAspProProIleGlyAlaGlu 240
              |||||||||||||||||||||||||||:::|||||||||:::||||||:::|||||||||
Db        690 GCCCTGCGCTTCCTGTGGGGCTCTGAGGAGGGCTGGGAAGAGCCCCCTGTTGGGGCTGAG 749

Qy        241 ---GlnSerPheHisLeuAlaLeuLeuArgGlnProGlyIleTyrLysProPheIleIle 259
                 |||   |||   ||||||:::||||||:::||||||::::::||||||   ||||||
Db        750 CACCAGGGCTTCCAGCTGGCTATGCTGAGGCGCCCTGGTGTCCACAAGCCCCTCATCATC 809

Qy        260 GlyValSerLeuMetAlaPheGlnGlnLeuSerGlyValAsnAlaValMetPheTyrAla 279
              |||:::   ||||||   |||||||||||||||||||||||||||:::||||||||||||
Db        810 GGCATTTGCCTCATGGTCTTCCAGCAGCTGTCAGGGGTCAACGCCATCATGTTCTATGCC 869

Qy        280 GluThrIlePheGluGluAlaLysPheLysAspSerSerLeuAlaSerValValValGly 299
                 ||||||||||||||||||||||||||||||||||||||||||||||||   ||||||
Db        870 AACACCATCTTTGAGGAGGCCAAGTTCAAGGACAGCAGCCTGGCCTCGGTCACTGTGGGC 929

Qy        300 ValIleGlnValLeuPheThrAlaValAlaAlaLeuIleMetAspArgAlaGlyArgArg 319
              :::||||||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db        930 ATCATCCAGGTTCTGTTCACTGCTGTGGCGGCCCTCATCATGGACAGAGCAGGGCGAAAA 989

Qy        320 LeuLeuLeuValLeuSerGlyValValMetValPheSerThrSerAlaPheGlyAlaTyr 339
              |||||||||   ||||||||||||:::||||||||||||   ||||||||||||   |||
Db        990 CTGCTTCTGGCCTTGTCGGGTGTGATCATGGTTTTCAGCATGAGCGCCTTTGGTACCTAC 1049

Qy        340 PheLysLeuThrGlnGlyGlyProGlyAsnSerSerHisValAlaIleSerAlaProVal 359
              |||||||||||||||   ||||||   |||||||||||||||   :::      |||:::
Db       1050 TTCAAACTGACCCAGAGTGGCCCCAGCAACTCCTCCCATGTAGGCCTCCTGGTGCCCATC 1109

Qy        360 SerAlaGlnProValAspAlaSerValGlyLeuAlaTrpLeuAlaValGlySerMetCys 379
              ||||||:::|||   |||      :::||||||||||||||||||||||||||||||
Db       1110 TCCGCAGAGCCTGCTGATGTTCACCTGGGGCTGGCCTGGCTGGCTGTAGGCAGCATGTGC 1169

Qy        380 LeuPheIleAlaGlyPheAlaValGlyTrpGlyProIleProTrpLeuLeuMetSerGlu 399
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1170 CTCTTCATCGCTGGTTTTGCAGTAGGCTGGGGACCCATCCCCTGGCTCCTCATGTCAGAG 1229

Qy        400 IlePheProLeuHisValLysGlyValAlaThrGlyIleCysValLeuThrAsnTrpLeu 419
              |||||||||||||||:::||||||||||||||||||:::||||||||||||||||||
Db       1230 ATCTTCCCTCTGCACATCAAGGGTGTGGCTACCGGCGTCTGTGTCCTCACCAACTGGTTC 1289

Qy        420 MetAlaPheLeuValThrLysGluPheSerSerLeuMetGluValLeuArgProTyrGly 439
              |||||||||||||||||||||||||||:::|||:::||||||:::|||||||||||||||
Db       1290 ATGGCCTTTCTGGTGACCAAAGAGTTTAACAGCATCATGGAGATCCTCAGACCCTACGGC 1349

Qy        440 AlaPheTrpLeuAlaSerAlaPheCysIlePheSerValLeuPheThrLeuPheCysVal 459
              ||||||||||||   :::||||||||||||   ||||||||||||||||||      |||
Db       1350 GCCTTCTGGCTCACCGCTGCCTTCTGTATCCTCAGCGTCCTTTTCACGCTCACCTTTGTC 1409

Qy        460 ProGluThrLysGlyLysThrLeuGluGlnIleThrAlaHisPheGluGlyArg 477
              |||||||||||||||:::||||||||||||||||||||||||||||||||||||
Db       1410 CCTGAGACTAAAGGCAGGACTCTGGAACAAATCACAGCCCATTTCGAGGGACGG 1463

RESULT 7
AAF55867
ID   AAF55867 standard; cDNA; 2072 BP.
XX
AC   AAF55867;
XX
DT   17-APR-2001  (first entry)
XX
DE   Murine GLUTX1 coding sequence.
XX
KW   Murine; GLUTX; gene therapy; vaccine; hexose transport modulator;
KW   hexose transport disorder; ischaemia; diabetes; hyperglycaemia; ss;
KW   hypoglycaemia; glucose metabolism disorder; neurodegenerative disease.
XX
OS   Mus sp.
XX
PN   WO200104145-A2.
XX
PD   18-JAN-2001.
XX
PF   14-JUL-2000; 2000WO-IB001042.
XX
PR   14-JUL-1999;    99US-0143907P.
PR   27-AUG-1999;    99US-0151140P.
PR   23-FEB-2000; 2000US-0184285P.
PR   13-JUL-2000; 2000US-00616132.
XX
PA   (UYLA-) UNIV LAUSANNE.
XX
PI   Thorens B,  Ibberson M,  Uldry M;
XX
DR   WPI; 2001-112615/12.
DR   P-PSDB; AAB66934.
XX
PT   Nucleic acids encoding GLUTX glucose transporter proteins, useful in the
PT   prevention, diagnosis and treatment of hexose transport disorders, e.g.
```

```
PT   ischemia and diabetes.
XX
PS   Claim 3; Page 73-74; 124pp; English.
XX
CC   The present invention relates to GLUTX proteins (AAF55865-AAF55871 and
CC   AAB66932-AAB66941). The GLUTX proteins are related to the facultative
CC   glucose carriers GLUT1-GLUT5 and have hexose binding and/or transport
CC   function. The GLUTX proteins may be used in the diagnosis, prevention and
CC   treatment of hexose transport disorders such as ischaemia, diabetes,
CC   hyperglycaemia, hypoglycaemia, a glucose metabolism disorder and/or a
CC   neurodegenerative disease. The present sequence is the coding sequence
CC   for murine GLUTX1
XX
SQ   Sequence 2072 BP; 351 A; 673 C; 584 G; 464 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              1.02e-177      Length:         2072
Score:                  2150.00        Matches:        409
Percent Similarity:     92.47%         Conservative:   33
Best Local Similarity:  85.56%         Mismatches:     34
Query Match:            87.51%         Indels:         2
DB:                     4              Gaps:           2

US-09-886-954A-1 (1-477) x AAF55867 (1-2072)

Qy          1 MetThrProGluAspProGluGluThrGlnProLeuLeuGlyProProGlyGlySerAla 20
              |||:::||||||||||||:::||||||||||||||||||   ||||||
Db         21 ATGTCTCCCGAGGACCCCCAGGAGACGCAGCCGCTATTGCGGCCACCGGAAGCCAGGACT 80

Qy         21 ProArgGlyArgArgValPheLeuAlaAlaPheAlaAlaAlaLeuGlyProLeuSerPhe 40
              |||||||||||||||||||||||||||:::||||||||||||||||||||||||:::|||
Db         81 CCCCGCGGCCGCCGGGTCTTCCTCGCTTCCTTCGCCGCCGCTCTGGGACCCCTCAACTTC 140

Qy         41 GlyPheAlaLeuGlyTyrSerSerProAlaIleProSerLeuGlnArgAlaAlaProPro 60
              ||||||||||||||||||||||||||||||||||||||||||:::|||   |||||||||
Db        141 GGCTTCGCGCTCGGCTACAGCTCCCCCGCCATCCCCAGCCTGCGGCGCACCGCACCCCCG 200

Qy         61 AlaProArgLeuAspAspAlaAlaAlaSerTrpPheGlyAlaValValThrLeuGlyAla 80
              |||   ||||||   |||   |||||||||||||||||||||||||||||||||||||||
Db        201 GCCCTGCGCCTCGGAGACAATGCGGCCTCCTGGTTCGGGGCCGTCGTGACCCTGGGCGCT 260

Qy         81 AlaAlaGlyGlyValLeuGlyGlyTrpLeuValAspArgAlaGlyArgLysLeuSerLeu 100
              ||||||||||||:::|||||||||||||||:::||||||:::||||||||||||||||||
Db        261 GCTGCAGGGGGCATACTGGGCGGCTGGCTCCTGGACCGTTCAGGGCGCAAGCTGAGCCTC 320

Qy        101 LeuLeuCysSerValProPheValAlaGlyPheAlaValIleThrAlaAlaGlnAspVal 120
              |||||||||:::||||||||||||   ||||||||||||||||||||||||:::||||||
Db        321 TTGCTCTGCACCGTGCCCTTCGTGACTGGCTTTGCTGTCATCACCGCGGCCCGGGATGTG 380

Qy        121 TrpMetLeuLeuGlyGlyArgLeuLeuThrGlyLeuAlaCysGlyValAlaSerLeuVal 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        381 TGGATGCTGCTCGGAGGCCGCCTCCTCACCGGCCTTGCCTGCGGAGTCGCCTCACTAGTG 440

Qy        141 AlaProValTyrIleSerGluIleAlaTyrProAlaValArgGlyLeuLeuGlySerCys 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        441 GCACCGGTTTACATCTCGGAAATCGCCTACCCAGCTGTCCGAGGACTGCTCGGCTCCTGT 500

Qy        161 ValGlnLeuMetValValValGlyIleLeuLeuAlaTyrLeuAlaGlyTrpValLeuGlu 180
              ||||||||||||||||||   ||||||||||||||||||:::||||||||||||||||||
Db        501 GTGCAGCTGATGGTTGTCACTGGCATCCTCCTGGCCTATGTGGCAGGCTGGGTCCTAGAG 560

Qy        181 TrpArgTrpLeuAlaValLeuGlyCysValProProSerLeuMetLeuLeuLeuMetCys 200
              ||||||||||||||||||||||||||||||||||||:::|||||||||||||||||||||
Db        561 TGGCGCTGGCTGGCCGTGCTGGGCTGTGTGCCCCCCACCCTCATGCTGCTGCTCATGTGC 620

Qy        201 PheMetProGluThrProArgPheLeuLeuThrGlnHisArgArgGlnGluAlaMetAla 220
              :::||||||||||||||||||||||||||||||||||||:::   |||||||||||||||
Db        621 TACATGCCCGAGACCCCACGTTTTCTCCTCACTCAACACCAGTACCAGGAGGCCATGGCT 680

Qy        221 AlaLeuArgPheLeuTrpGlySerGluGlnGlyTrpGluAspProProIleGlyAlaGlu 240
              |||||||||||||||||||||||||||:::|||||||||:::||||||:::|||||||||
Db        681 GCCTTGCGCTTCCTGTGGGGCTCTGAGGAGGGCTGGGAAGAGCCCCCTGTTGGGGCTGAG 740

Qy        241 ---GlnSerPheHisLeuAlaLeuLeuArgGlnProGlyIleTyrLysProPheIleIle 259
                 |||   |||   |||||||||||||||:::||||||||||||||||||   ||||||
Db        741 CACCAGGGCTTCCAGCTGGCCCTGCTGAGGCGCCCTGGCATCTACAAGCCCCTCATCATC 800

Qy        260 GlyValSerLeuMetAlaPheGlnGlnLeuSerGlyValAsnAlaValMetPheTyrAla 279
              |||:::|||||||||   |||||||||||||||||||||||||||:::||||||||||||
Db        801 GGCATTTCCCTCATGGTCTTCCAGCAGCTGTCAGGGGTCAATGCTATCATGTTCTATGCC 860

Qy        280 GluThrIlePheGluGluAlaLysPheLysAspSerSerLeuAlaSerValValValGly 299
                 :::|||||||||||||||||||||||||||||||||||||||||||||   ||||||
Db        861 AACAGCATCTTCGAGGAGGCCAAGTTCAAGGACAGCAGCCTGGCCTCGGTCACTGTGGGC 920

Qy        300 ValIleGlnValLeuPheThrAlaValAlaAlaLeuIleMetAspArgAlaGlyArgArg 319
              :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        921 ATAATCCAGGTCCTGTTCACTGCTGTGGCGGCCCTCATCATGGACAGAGCAGGGCGAAGG 980

Qy        320 LeuLeuLeuValLeuSerGlyValValMetValPheSerThrSerAlaPheGlyAlaTyr 339
              |||||||||   ||||||||||||:::||||||||||||   ||||||||||||   |||
Db        981 CTGCTCCTGGCCTTGTCGGGTGTGATCATGGTGTTCAGTATGAGTGCCTTTGGTACCTAC 1040

Qy        340 PheLysLeuThrGlnGlyGlyProGlyAsnSerSerHisValAlaIleSerAlaProVal 359
              |||||||||||||||      |||   |||||||||||||||   :::      |||:::
Db       1041 TTCAAACTGACCCAGAGCCTCCCCAGCAACTCCTCCCACGTAGGCCTG---GTGCCCATC 1097

Qy        360 SerAlaGlnProValAspAlaSerValGlyLeuAlaTrpLeuAlaValGlySerMetCys 379
              :::|||:::|||||||||      ||||||||||||||||||||||||||||||||||||
Db       1098 GCGGCGGAGCCTGTGGATGTCCAAGTGGGACTGGCCTGGCTGGCTGTAGGCAGCATGTGC 1157

Qy        380 LeuPheIleAlaGlyPheAlaValGlyTrpGlyProIleProTrpLeuLeuMetSerGlu 399
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1158 CTCTTCATTGCTGGCTTTGCGGTGGGCTGGGGACCCATCCCCTGGCTCCTCATGTCAGAG 1217

Qy        400 IlePheProLeuHisValLysGlyValAlaThrGlyIleCysValLeuThrAsnTrpLeu 419
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1218 ATCTTCCCTCTGCATGTCAAGGGTGTGGCTACCGGCATCTGTGTCCTCACCAACTGGTTC 1277

Qy        420 MetAlaPheLeuValThrLysGluPheSerSerLeuMetGluValLeuArgProTyrGly 439
              |||||||||||||||||||||||||||:::|||:::||||||:::|||||||||||||||
Db       1278 ATGGCCTTTCTAGTGACCAAAGAGTTCAACAGCGTCATGGAGATGCTCAGACCCTACGGT 1337

Qy        440 AlaPheTrpLeuAlaSerAlaPheCysIlePheSerValLeuPheThrLeuPheCysVal 459
              ||||||||||||   :::|||||||||      ||||||||||||||||||      |||
Db       1338 GCCTTCTGGCTCACCGCTGCCTTCTGCGCTCTCAGTGTCCTATTCACACTGACCGTTGTC 1397

Qy        460 ProGluThrLysGlyLysThrLeuGluGlnIleThrAlaHisPheGluGlyArg 477
              |||||||||||||||:::||||||||||||:::|||||||||||||||||||||
Db       1398 CCTGAGACTAAAGGCAGGACTCTGGAACAAGTCACAGCCCATTTCGAGGGACGA 1451

RESULT 8
ABX34029
ID   ABX34029 standard; cDNA; 1954 BP.
XX
AC   ABX34029;
XX
DT   10-FEB-2003  (first entry)
XX
DE   Human cancer suppressing protein PP7425.
XX
KW   Human; gene; ss; cancer suppressing protein; cancer.
XX
OS   Homo sapiens.
XX
PN   CN1351081-A.
XX
PD   29-MAY-2002.
XX
PF   31-OCT-2000; 2000CN-00127102.
XX
PR   31-OCT-2000; 2000CN-00127102.
XX
PA   (SHAN-) SHANGHAI INST ONCOLOGY.
XX
PI   Gu J;
```

```
XX
DR   WPI; 2002-609437/66.
DR   P-PSDB; ABU11283.
XX
PT   New human protein with cancer cell growth suppressing function and a
PT   polynucleotide encoding it, for treating diseases, such as, cancer.
XX
PS   Claim 5; Page 23-24 (disclosure); 39pp; Chinese.
XX
CC   This invention relates to the cDNA and protein sequences of a novel human
CC   protein with cancer suppressing function. The invention also comprises a
CC   method for preparing the polypeptide by recombination, and an application
CC   of the polypeptide in treating diseases such as cancer, etc. Also
CC   disclosed in an antagonist of the polypeptide and its medical action. The
CC   present sequence represents a cDNA encoding a cancer suppressing protein
CC   of the invention
XX
SQ   Sequence 1954 BP; 325 A; 633 C; 573 G; 423 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              1.94e-175    Length:         1954
Score:                  2123.50      Matches:        426
Percent Similarity:     89.14%       Conservative:   1
Best Local Similarity:  88.94%       Mismatches:     1
Query Match:            86.43%       Indels:         51
DB:                     6            Gaps:           1

US-09-886-954A-1 (1-477) x ABX34029 (1-1954)

Qy          1 MetThrProGluAspProGluGluThrGlnProLeuLeuGlyProProGlyGlySerAla 20
              :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTGACGCCCGAGGACCCAGAGGAAACCCAGCCGCTTCTGGGGCCTCCTGGCGGCAGCGCG 60

Qy         21 ProArgGlyArgArgValPheLeuAlaAlaPheAlaAlaAlaLeuGlyProLeuSerPhe 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCCCGCGGCCGCCGCGTCTTCCTCGCCGCCTTCGCCGCTGCCCTGGGCCCACTCAGCTTC 120

Qy         41 GlyPheAlaLeuGlyTyrSerSerProAlaIleProSerLeuGlnArgAlaAlaProPro 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGCTTCGCGCTCGGCTACAGCTCCCCGGCCATCCCTAGCCTGCAGCGCGCCGCGCCCCCG 180

Qy         61 AlaProArgLeuAspAspAlaAlaAlaSerTrpPheGlyAlaValValThrLeuGlyAla 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GCCCCGCGCCTGGACGACGCCGCCGCCTCCTGGTTCGGGGCTGTCGTGACCCTGGGTGCC 240

Qy         81 AlaAlaGlyGlyValLeuGlyGlyTrpLeuValAspArgAlaGlyArgLysLeuSerLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCGGCGGGGGGAGTGCTGGGCGGCTGGCTGGTGGACCGCGCCGGGCGCAAGCTGAGCCTC 300

Qy        101 LeuLeuCysSerValProPheValAlaGlyPheAlaValIleThrAlaAlaGlnAspVal 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TTGCTGTGCTCCGTGCCCTTCGTGGCCGGCTTTGCCGTCATCACCGCGGCCCAGGACGTG 360

Qy        121 TrpMetLeuLeuGlyGlyArgLeuLeuThrGlyLeuAlaCysGlyValAlaSerLeuVal 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TGGATGCTGCTGGGGGGCCGCCTCCTCACCGGCCTGGCCTGCGGTGTTGCCTCCCTAGTG 420

Qy        141 AlaProValTyrIleSerGluIleAlaTyrProAlaVal-ArgGlyLeuLeuGlySerCy 160
              ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db        421 GCCCCGGTCTACATCTCCGAAATCGCCTACCCAGCAGTCCCGGGGGTTGCTCGGCTCCTG 480

Qy        160 sValGlnLeuMetValValValGlyIleLeuLeuAlaTyrLeuAlaGlyTrpValLeuGl 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TGTGCAGCTAATGGTCGTCGTCGGCATCCTCCTGGCCTACCTGGCAGGCTGGGTGCTGGA 540

Qy        180 uTrpArgTrpLeuAlaValLeuGlyCysValProProSerLeuMetLeuLeuLeuMetCy 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTGGCGCTGGCTGGCTGTGCTGGGCTGCGTGCCCCCCTCCCTCATGCTGCTTCTCATGTG 600

Qy        200 sPheMetProGluThrProArgPheLeuLeuThrGlnHisArgArgGlnGluAlaMetAl 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTTCATGCCCGAGACCCCGCGCTTCCTGCTGACTCAGCACAGGCGCCAGGAGGCCATGGC 660

Qy        220 aAlaLeuArgPheLeuTrpGlySerGluGlnGlyTrpGluAspProProIleGlyAlaGl 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CGCCCTGCGGTTCCTGTGGGGCTCCGAGCAGGGCTGGGAAGACCCCCCCATCGGGGCTGA 720

Qy        240 uGlnSerPheHisLeuAlaLeuLeuArgGlnProGlyIleTyrLysProPheIleIleGl 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GCAGAGCTTTCACCTGGCCCTGCTGCGGCAGCCCGGCATCTACAAGCCCTTCATCATCGG 780

Qy        260 yValSerLeuMetAlaPheGlnGlnLeuSerGlyValAsnAlaValMetPheTyrAlaGl 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CGTCTCCCTGATGGCCTTCCAGCAGCTGTCGGGGGTCAACGCCGTCATGTTCTATGCAGA 840

Qy        280 uThrIlePheGluGluAlaLysPheLysAspSerSerLeuAlaSerValValValGlyVa 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GACCATCTTTGAAGAGGCCAAGTTCAAGGACAGCAGCCTGGCCTCGGTCGTCGTGGGTGT 900

Qy        300 lIleGlnValLeuPheThrAlaValAlaAlaLeuIleMetAspArgAlaGlyArgArgLe 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 CATCCAGGTGCTGTTCACAGCTGTGGCGGCTCTCATCATGGACAGAGCAGGGCGGAGGCT 960

Qy        320 uLeuLeuValLeuSerGlyValValMetValPheSerThrSerAlaPheGlyAlaTyrPh 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GCTCCTGGTCTTGTCAGGTGTGGTCATGGTGTTCAGCACGAGTGCCTTCGGCGCCTACTT 1020

Qy        340 eLysLeuThrGlnGlyGlyProGlyAsnSerSerHisValAlaIleSerAlaProValSe 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CAAGCTGACCCAGGGTGGCCCTGGCAACTCCTCGCACGTGGCCATCTCGGCGCCTGTCTC 1080

Qy        360 rAlaGlnProValAspAlaSerValGlyLeuAlaTrpLeuAlaValGlySerMetCysLe 380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 TGCACAGCCTGTTGATGCCAGCGTGGGGCTGGCCTGGCTGGCCGTGGGCAGCATGTGCCT 1140

Qy        380 uPheIleAlaGly-PheAlaValGlyTrpGlyProIleProTrpLeuLeuMetSerGluI 400
              |||||||||||||
Db       1141 CTTCATCGCCGGAGGTC------------------------------------------- 1157

Qy        400 lePheProLeuHisValLysGlyValAlaThrGlyIleCysValLeuThrAsnTrpLeuM 420

Db       1157 ------------------------------------------------------------ 1157

Qy        420 etAlaPheLeuValThrLysGluPheSerSerLeuMetGluValLeuArgProTyrGlyA 440
                                                          ||||||||||||||||
Db       1158 --------------------------------------------CTCAGGCCCTATGGAG 1173

Qy        440 laPheTrpLeuAlaSerAlaPheCysIlePheSerValLeuPheThrLeuPheCysValP 460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1174 CCTTCTGGCTTGCCTCCGCTTTCTGCATCTTCAGTGTCCTTTTCACTTTGTTCTGTGTCC 1233

Qy        460 roGluThrLysGlyLysThrLeuGluGlnIleThrAlaHisPheGluGlyArg 477
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1234 CTGAAACTAAAGGAAAGACTCTGGAACAAATCACAGCCCATTTTGAGGGGCGA 1286

RESULT 9
ABL90126
ID   ABL90126 standard; cDNA; 1156 BP.
XX
AC   ABL90126;
XX
DT   24-MAY-2002  (first entry)
XX
DE   Human polynucleotide SEQ ID NO 688.
XX
KW   Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW   antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
KW   vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW   cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW   neurological disease; infection; human; secreted protein; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200190304-A2.
XX
PD   29-NOV-2001.
```

```
XX
PF   18-MAY-2001; 2001WO-US016450.
XX
PR   19-MAY-2000; 2000US-0205515P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Birse CE,  Rosen CA;
XX
DR   WPI; 2002-122018/16.
DR   P-PSDB; ABB89717.
XX
PT   Novel 1405 isolated polypeptides, useful for diagnosis, treatment and
PT   prevention of neural, immune system, muscular, reproductive,
PT   gastrointestinal, pulmonary, cardiovascular, renal and proliferative
PT   disorders.
XX
PS   Claim 4; SEQ ID NO 688; 2081pp + Sequence Listing; English.
XX
CC   The invention relates to novel genes (ABL89449-ABL90853) and proteins
CC   (ABB89040-ABB90444) useful for preventing, treating or ameliorating
CC   medical conditions e.g. by protein or gene therapy. The genes are
CC   isolated from a range of human tissues disclosed in the specification.
CC   The nucleic acids, proteins, antibodies and (ant)agonists are useful in
CC   the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
CC   ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,
CC   breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune
CC   disorders e.g. Addison's disease, allergies, autoimmune haemolytic
CC   anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
CC   multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
CC   cardiovascular disorders such as myocardial ischaemias; (d) wound healing
CC   ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
CC   infectious diseases such as viral, bacterial, fungal and parasitic
CC   infections. Note: The sequence data for this patent did not form part of
CC   the printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 1156 BP; 177 A; 362 C; 345 G; 255 T; 0 U; 17 Other;

Alignment Scores:
Pred. No.:              5.45e-137     Length:         1156
Score:                  1679.50       Matches:        344
Percent Similarity:     72.12%        Conservative:   0
Best Local Similarity:  72.12%        Mismatches:     10
Query Match:            68.36%        Indels:         125
DB:                     6             Gaps:           1

US-09-886-954A-1 (1-477) x ABL90126 (1-1156)

Qy          1 MetThrProGluAspProGluGluThrGlnProLeuLeuGlyProProGlyGlySerAla 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         53 ATGACGCCCGAGGACCCAGAGGAAACCCAGCCGCTTCTGGGGCCTCCTGGCGGCAR---- 108

Qy         21 ProArgGlyArgArgValPheLeuAlaAlaPheAlaAlaAlaLeuGlyProLeuSerPhe 40

Db        108 ------------------------------------------------------------ 108

Qy         41 GlyPheAlaLeuGlyTyrSerSerProAlaIleProSerLeuGlnArgAlaAlaProPro 60

Db        108 ------------------------------------------------------------ 108

Qy         61 AlaProArgLeuAspAspAlaAlaAlaSerTrpPheGlyAlaValValThrLeuGlyAla 80

Db        108 ------------------------------------------------------------ 108

Qy         81 AlaAlaGlyGlyValLeuGlyGlyTrpLeuValAspArgAlaGlyArgLysLeuSerLeu 100

Db        108 ------------------------------------------------------------ 108

Qy        101 LeuLeuCysSerValProPheValAlaGlyPheAlaValIleThrAlaAlaGlnAspVal 120

Db        108 ------------------------------------------------------------ 108
```

```
Qy        121 TrpMetLeuLeuGlyGlyArgLeuLeuThrGlyLeuAlaCysGlyValAlaSerLeuVal 140

Db        108 ------------------------------------------------------------ 108

Qy        141 AlaProValTyrIleSerGluIleAlaTyrProAlaValArgGlyLeuLeuGlySerCys 160
                    ||||||||||||||||||   |||||||||||||||||||||||||||||||||
Db        109 ------GTCTACATCTCCGAAATCRCCTACCCAGCAGTCCGGGGGTTGCTCGGCTCCTGT 162

Qy        161 ValGlnLeuMetValValValGlyIleLeuLeuAlaTyrLeuAlaGlyTrpValLeuGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        163 GTGCAGCTAATGGTCGTCGTCGGCATCCTCCTGGCCTACCTGGCAGGCTGGGTGCTGGAG 222

Qy        181 TrpArgTrpLeuAlaValLeuGlyCysValProProSerLeuMetLeuLeuLeuMetCys 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        223 TGGCGCTGGCTGGCTGTGCTGGGCTGCGTGCCCCCCTCCCTCATGCTGCTTCTCATGTGC 282

Qy        201 PheMetProGluThrProArgPheLeuLeuThrGlnHisArgArgGlnGluAlaMetAla 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        283 TTCATGCCCGAGACCCCGCGCTTCCTGCTGACTCAGCACAGGCGCCAGGAGGCCATGGCC 342

Qy        221 AlaLeuArgPheLeuTrpGlySerGluGlnGlyTrpGluAspProProIleGlyAlaGlu 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||   |||
Db        343 GCCCTGCGNTTCCTGTGGGGCTCCGAGCAGGGCTGGAAGACCCCCCCATCGGGNCTGAG 402

Qy        241 GlnSerPheHisLeuAlaLeuLeuArgGlnProGlyIleTyrLysProPheIleIleGly 260
              |||||||||||||||||||||||||||   ||||||||||||||||||||||||||||||
Db        403 CAGAGCTTTCACCTGGCCCTGCTGCGGCANCCCGGCATCTACAAGCCCTTCATCATCGGC 462

Qy        261 ValSerLeuMetAlaPheGlnGlnLeuSerGlyValAsnAlaValMetPheTyrAlaGlu 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        463 GTCTCCCTGATGGCCTTCCAGCAGCTGTCGGGGGTCAACGCCGTCATGTTCTATGCAGAG 522

Qy        281 ThrIlePheGluGluAlaLysPheLysAspSerSerLeuAlaSerValValValGlyVal 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        523 ACCATCTTTGAAGAGGCCAAGTTCAAGGACAGCAGCCTGGCCTCGGTCGTCGTGGGTGTC 582

Qy        301 IleGlnValLeuPheThrAlaValAlaAlaLeuIleMetAspArgAlaGlyArgArgLeu 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        583 ATCCAGGTGCTGTTCACAGCTGTGGCGGCTCTCATCATGGACAGAGCAGGGCGGAGGCTG 642

Qy        321 LeuLeuValLeuSerGlyValValMetValPheSerThrSerAlaPheGlyAlaTyrPhe 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        643 CTCCTGGTCTTGTCAGGTGTGGTCATGGTGTTCAGCACGAGTGCCTTCGGCGCCTACTTC 702

Qy        341 LysLeuThrGlnGlyGlyProGlyAsnSerSerHisValAlaIleSerAlaProValSer 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        703 AAGCTGACCCAGGGTGGCCCTGGCAACTCCTCGCACGTGGCCATCTCGGCGCCTGTCTCT 762

Qy        361 AlaGlnProValAspAlaSerValGlyLeuAlaTrpLeuAlaValGlySerMetCysLeu 380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        763 GCACAGCCTGTTGATGCCAGCGTGGGGCTNGCCTGGCTGGCCGTGGGCAGCATGTGCCTC 822

Qy        381 PheIleAlaGlyPheAlaValGlyTrpGlyProIleProTrpLeuLeuMetSerGluIle 400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        823 TTCATCGCCGGCTTTGCGGTGGGCTGGGGGCCCATCCCCTGGCTCCTCATGTCAGAGATC 882

Qy        401 PheProLeuHisValLysGlyValAlaThrGlyIleCysValLeuThrAsnTrpLeuMet 420
              |||||||||||||||||||||||||||||||||   |||||||||   ||||||||||||
Db        883 TTCCCTYTGCATGTCAAGGGCGTGGCGACAGGCAWYTGCGTCCTCAMCAACTGGCTCATG 942

Qy        421 AlaPheLeuValThrLysGluPheSerSerLeuMetGluValLeuArgProTyrGlyAla 440
                 ||||||   ||||||||||||||||||||||||||||||||||||||||||||||||
Db        943 GSCTTTCTYGKGACCAAGGAGTTCAGCAGCCTCATGGAGGTCCTCAGGCCCTATGGAGCC 1002

Qy        441 PheTrpLeuAlaSerAlaPheCysIlePheSerValLeuPheThrLeuPheCysValPro 460
              ||||||||||||||||||||||||||||||||||||||||||||||||   |||   |||
Db       1003 TTCTGGCTTGCCTCCGCTTTCTGCATCTTCAGTGTCCTTTTCACTTTGGTCTGTGC-CCT 1061

Qy        461 GluThrLysGlyLysThrLeuGluGlnIleThrAlaHisPheGluGlyArg 477
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1062 GAAACTAAAGGAAAGACTCTGGAACAAATCACAGCCCATTTTGAGGGGCGA 1112

RESULT 10
```

```
AAD12574
ID   AAD12574 standard; cDNA; 1461 BP.
XX
AC   AAD12574;
XX
DT   25-SEP-2001  (first entry)
XX
DE   Human protein having hydrophobic domain encoding cDNA clone HP10784.
XX
KW   Human; hydrophobic domain; gene therapy; nutritional supplement;
KW   cell proliferation; immunomodulatory; autoimmune disorder; antimicrobial;
KW   multiple sclerosis; rheumatoid arthritis; insulin-dependent diabetes;
KW   haematopoiesis; tissue growth activity; Parkinson's disease; cytostatic;
KW   Huntington's disease; Alzheimer's disease; chemotactic; chemokinetic;
KW   haemostatic; thrombolytic; tumour growth inhibitor; anabolic;
KW   contraceptive; antiinfertility; antiinflammatory; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             61. .849
FT                   /*tag=  a
FT                   /product= "Human protein having hydrophobic domain"
FT                   /note= "CDS is specifically is claimed in claim 3"
XX
PN   WO200149728-A2.
XX
PD   12-JUL-2001.
XX
PF   28-DEC-2000; 2000WO-JP009359.
XX
PR   06-JAN-2000; 2000JP-00000585.
PR   06-JAN-2000; 2000JP-00000588.
PR   11-JAN-2000; 2000JP-00002299.
PR   03-FEB-2000; 2000JP-00026862.
PR   03-MAR-2000; 2000JP-00058367.
XX
PA   (PROT-) PROTEGENE INC.
PA   (SAGA ) SAGAMI CHEM RES CENT.
XX
PI   Kato S,  Kimura T;
XX
DR   WPI; 2001-418355/44.
DR   P-PSDB; AAE06579.
XX
PT   Human proteins with hydrophobic domains and the nucleic acids encoding
PT   them, useful for preventing diagnosing and treating e.g. cancer,
PT   Alzheimer's and inflammation.
XX
PS   Claim 4; Page 287-289; 563pp; English.
XX
CC   The present sequence is human protein with hydrophobic domain encoding
CC   cDNA clone HP10784. The polynucleotide and polypeptide of the invention
CC   may be used in the prevention, diagnosis and treatment of diseases
CC   associated with inappropriate polypeptide expression. The polynucleotides
CC   may be used to produce the polypeptide, by inserting the nucleic acids
CC   into a host cell and culturing the cell to express the protein. The
CC   polynucleotides and its complementary sequences may also be used as DNA
CC   probes in diagnostic assays and also used in gene therapy. The
CC   polypeptides may also be used as antigens in the production of antibodies
CC   and in assays to identify modulators of polypeptide expression and
CC   activity. The polypeptides and nucleic acids may be used as nutritional
CC   supplements, to modulate cytokine and cell proliferation activity, to
CC   modulate immune stimulation or suppression (e.g. for the treatment of
CC   microbial infections and autoimmune disorders such as multiple sclerosis,
CC   rheumatoid arthritis and insulin-dependent diabetes), to modulate
CC   haematopoiesis, to modulate tissue growth activity (e.g. for the
CC   treatment of Parkinson's disease, Huntington's disease and Alzheimer's
CC   disease), to modulate activin and inhibin activity (e.g. for controlling
CC   fertility), to modulate chemotactic and chemokinetic activity, to
CC   modulate haemostatic and thrombolytic activity, to modulate receptor
CC   ligand activity, to modulate inflammation and to inhibit tumour growth
XX
```

```
SQ   Sequence 1461 BP; 195 A; 501 C; 456 G; 309 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              6.03e-126     Length:         1461
Score:                  1554.50       Matches:        324
Percent Similarity:     67.85%        Conservative:   1
Best Local Similarity:  67.64%        Mismatches:     1
Query Match:            63.27%        Indels:         153
DB:                     4             Gaps:           2

US-09-886-954A-1 (1-477) x AAD12574 (1-1461)

Qy          1 MetThrProGluAspProGluGluThrGlnProLeuLeuGlyProProGlyGlySerAla 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ATGACGCCCGAGGACCCAGAGGAAACCCAGCCGCTTCTGGGGCCTCCTGGCGGCAGCGCG 120

Qy         21 ProArgGlyArgArgValPheLeuAlaAlaPheAlaAlaAlaLeuGlyProLeuSerPhe 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCCCGCGGCCGCCGCGTCTTCCTCGCCGCCTTCGCCGCTGCCCTGGGCCCACTCAGCTTC 180

Qy         41 GlyPheAlaLeuGlyTyrSerSerProAlaIleProSerLeuGlnArgAlaAlaProPro 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCTTCGCGCTCGGCTACAGCTCCCCGGCCATCCCTAGCCTGCAGCGCGCCGCGCCCCCG 240

Qy         61 AlaProArgLeuAspAspAlaAlaAlaSerTrpPheGlyAlaValValThrLeuGlyAla 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCCCCGCGCCTGGACGACGCCGCCGCCTCCTGGTTCGGGGCTGTCGTGACCCTGGGTGCC 300

Qy         81 AlaAlaGlyGlyValLeuGlyGlyTrpLeuValAspArgAlaGlyArgLysLeuSerLeu 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCGGCGGGGGGAGTGCTGGGCGGCTGGCTGGTGGACCGCGCCGGGCGCAAGCTGAGCCTC 360

Qy        101 LeuLeuCysSerValProPheValAlaGlyPheAlaValIleThrAlaAlaGlnAspVal 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TTGCTGTGCTCCGTGCCCTTCGTGGCCGGCTTTGCCGTCATCACCGCGGCCCAGGACGTG 420

Qy        121 TrpMetLeuLeuGlyGlyArgLeuLeuThrGlyLeuAlaCysGlyValAlaSerLeuVal 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TGGATGCTGCTGGGGGGCCGCCTCCTCACCGGCCTGGCCTGCGGTGTTGCCTCCCTAGTG 480

Qy        141 AlaProValTyrIleSerGluIleAlaTyrProAlaValArgGlyLeuLeuGlySerCys 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GCCCCGGTCTACATCTCCGAAATCGCCTACCCAGCAGTCCGGGGGTTGCTCGGCTCCTGT 540

Qy        161 ValGlnLeuMetValValValGlyIleLeuLeuAlaTyrLeuAlaGlyTrpValLeuGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTGCAGCTAATGGTCGTCGTCGGCATCCTCCTGGCCTACCTGGCAGGCTGGGTGCTGGAG 600

Qy        181 TrpArgTrpLeuAlaValLeuGlyCysValProProSerLeuMetLeuLeuLeuMetCys 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TGGCGCTGGCTGGCTGTGCTGGGCTGCGTGCCCCCCTCCCTCATGCTGCTTCTCATGTGC 660

Qy        201 PheMetProGluThrProArgPheLeuLeuThrGlnHisArgArgGlnGluAla-MetAl 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||  :::
Db        661 TTCATGCCCGAGACCCCGCGCTTCCTGCTGACTCAGCACAGGCGCCAGGAGGCTGCTC-- 718

Qy        220 aAlaLeuArgPheLeuTrpGlySerGluGlnGlyTrpGluAspProProIleGlyAlaGl 240

Db        718 ------------------------------------------------------------ 718

Qy        240 uGlnSerPheHisLeuAlaLeuLeuArgGlnProGlyIleTyrLysProPheIleIleGl 260

Db        718 ------------------------------------------------------------ 718

Qy        260 yValSerLeuMetAlaPheGlnGlnLeuSerGlyValAsnAlaValMetPheTyrAlaGl 280

Db        718 ------------------------------------------------------------ 718

Qy        280 uThrIlePheGluGluAlaLysPheLysAspSerSerLeuAlaSerValValValGlyVa 300

Db        718 ------------------------------------------------------------ 718

Qy        300 lIleGlnValLeuPheThrAlaValAlaAlaLeuIleMetAspArgAlaGlyArgArgLe 320
```

```
Db          718 ------------------------------------------------------------ 718

Qy          320 uLeuLeuValLeuSerGlyValValMetValPheSerThrSerAlaPheGlyAlaTyrPh 340
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          719 ----CTGGTCTTGTCAGGTGTGGTCATGGTGTTCAGCACGAGTGCCTTCGGCGCCTACTT 774

Qy          340 eLysLeuThrGlnGlyGlyProGlyAsnSerSerHisValAlaIleSerAlaProValSe 360
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          775 CAAGCTGACCCAGGGTGGCCCTGGCAACTCCTCGCACGTGGCCATCTCGGCGCCTGTCTC 834

Qy          360 rAlaGlnProValAspAlaSerValGlyLeuAlaTrpLeuAlaValGlySerMetCysLe 380
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          835 TGCACAGCCTGTTGATGCCAGCGTGGGGCTGGCCTGGCTGGCCGTGGGCAGCATGTGCCT 894

Qy          380 uPheIleAlaGly-PheAlaValGlyTrpGlyProIleProTrpLeuLeuMetSerGluI 400
                |||||||||||||
Db          895 CTTCATCGCCGGAGGTC------------------------------------------- 911

Qy          400 lePheProLeuHisValLysGlyValAlaThrGlyIleCysValLeuThrAsnTrpLeuM 420

Db          911 ------------------------------------------------------------ 911

Qy          420 etAlaPheLeuValThrLysGluPheSerSerLeuMetGluValLeuArgProTyrGlyA 440
                                                            ||||||||||||||||
Db          912 --------------------------------------------CTCAGGCCCTATGGAG 927

Qy          440 laPheTrpLeuAlaSerAlaPheCysIlePheSerValLeuPheThrLeuPheCysValP 460
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          928 CCTTCTGGCTTGCCTCCGCTTTCTGCATCTTCAGTGTCCTTTTCACTTTGTTCTGTGTCC 987

Qy          460 roGluThrLysGlyLysThrLeuGluGlnIleThrAlaHisPheGluGlyArg 477
                |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          988 CTGAAACTAAAGGAAAGACTCTGGAACAAATCACAGCCCATTTTGAGGGGCGA 1040

RESULT 11
AAF55871
ID   AAF55871 standard; cDNA; 2011 BP.
XX
AC   AAF55871;
XX
DT   17-APR-2001  (first entry)
XX
DE   Rat GLUTX3 coding sequence.
XX
KW   Rat; GLUTX; gene therapy; vaccine; hexose transport modulator;
KW   hexose transport disorder; ischaemia; diabetes; hyperglycaemia; ss;
KW   hypoglycaemia; glucose metabolism disorder; neurodegenerative disease.
XX
OS   Rattus sp.
XX
PN   WO200104145-A2.
XX
PD   18-JAN-2001.
XX
PF   14-JUL-2000; 2000WO-IB001042.
XX
PR   14-JUL-1999;   99US-0143907P.
PR   27-AUG-1999;   99US-0151140P.
PR   23-FEB-2000; 2000US-0184285P.
PR   13-JUL-2000; 2000US-00616132.
XX
PA   (UYLA-) UNIV LAUSANNE.
XX
PI   Thorens B,  Ibberson M,  Uldry M;
XX
DR   WPI; 2001-112615/12.
DR   P-PSDB; AAB66938.
XX
PT   Nucleic acids encoding GLUTX glucose transporter proteins, useful in the
PT   prevention, diagnosis and treatment of hexose transport disorders, e.g.
PT   ischemia and diabetes.
XX
```

```
PS   Claim 3; Page 82-83; 124pp; English.
XX
CC   The present invention relates to GLUTX proteins (AAF55865-AAF55871 and
CC   AAB66932-AAB66941). The GLUTX proteins are related to the facultative
CC   glucose carriers GLUT1-GLUT5 and have hexose binding and/or transport
CC   function. The GLUTX proteins may be used in the diagnosis, prevention and
CC   treatment of hexose transport disorders such as ischaemia, diabetes,
CC   hyperglycaemia, hypoglycaemia, a glucose metabolism disorder and/or a
CC   neurodegenerative disease. The present sequence is the coding sequence
CC   for rat GLUTX3
XX
SQ   Sequence 2011 BP; 337 A; 621 C; 593 G; 459 T; 0 U; 1 Other;

Alignment Scores:
Pred. No.:              5.75e-75        Length:         2011
Score:                  972.50          Matches:        219
Percent Similarity:     61.19%          Conservative:   79
Best Local Similarity:  44.97%          Mismatches:     169
Query Match:            39.58%          Indels:         21
DB:                     4               Gaps:           7

US-09-886-954A-1 (1-477) x AAF55871 (1-2011)

Qy            3 ProGluAspProGluGluThrGlnProLeuLeuGlyProProGlyGlySerAlaProArg 22
                |||      |||||||||                                 |||   :::
Db           64 CCCGCGTCGCCTGAAGAG------------------AAGGCACGGGCCGGGGCCTTGCAA 105

Qy           23 GlyArgArgValPheLeuAlaAlaPheAlaAlaAlaLeuGlyProLeuSerPheGlyPhe 42
                   ||||||||||||||||||   |||||||||   ||||||      |||||||||:::
Db          106 AACAGAAGGGTGTTCCTGGCCACCTTCGCTGCCGTGCTGGGCAATTTCAGCTTTGGGTAT 165

Qy           43 AlaLeuGlyTyrSerSerProAlaIleProSerLeuGlnArgAlaAlaProProAlaPro 62
                ||||||   |||:::||||||   ||||||:::|||:::|||::::::   ||||||
Db          166 GCCCTGGTCTACACATCTCCAGTCATCCCTGCGCTGAAGCGCTCTTCTGACCCAGCACTA 225

Qy           63 ArgLeuAspAspAlaAlaAlaSerTrpPheGlyAlaValValThrLeuGlyAlaAlaAla 82
                |||||||||         |||||||||||||||:::|||   ||||||||||||||||||
Db          226 CGCCTGGACAAAATCCAGGCATCCTGGTTTGGGTCCGTGTTCACCCTGGGTGCCGCCGCT 285

Qy           83 GlyGlyValLeuGlyGlyTrpLeuValAspArgAlaGlyArgLysLeuSerLeuLeuLeu 102
                ||||||:::            |||   |||      |||||||||||||||::::::
Db          286 GGGGGCCTCAGTGCTATGCTACTCAATGACCTCTTGGGCCGGAAGCTCAGCATCATGTTT 345

Qy          103 CysSerValProPheValAlaGlyPheAlaValIleThrAlaAlaGlnAspValTrpMet 122
                   :::||||||         |||:::|||::::::      |||:::   :::||||||
Db          346 TCCGCTGTCCCCTCGGCCATTGGCTATGCACTCATGGCTGGTGCCCGTGGCCTCTGGATG 405

Qy          123 LeuLeuGlyGlyArgLeuLeuThrGlyLeuAlaCysGlyValAlaSerLeuValAlaPro 142
                ||||||   ||||||:::|||||||||   |||   |||:::   :::         |||
Db          406 CTTCTGCTGGGGAGGATGCTGACAGGCTTTGCCGGGGGACTCACTGCTGCCTGCATCCCG 465

Qy          143 ValTyrIleSerGluIleAlaTyrProAlaValArgGlyLeuLeuGlySerCysValGln 162
                ||||||:::||||||||||||   |||   |||||||||   ||||||:::      |||
Db          466 GTGTACGTGTCTGAGATTGCACCCCCGGGTGTTCGCGGGGCCCTGGGGGCCACGCCGCAG 525

Qy          163 LeuMetValValValGlyIleLeuLeuAlaTyrLeuAlaGlyTrpValLeuGluTrpArg 182
                ||||||   |||   |||   |||      |||      |||   :::|||   ||||||
Db          526 CTCATGGCCGTGTTTGGATCCCTGTCTCTCTATGCCCTTGGTCTTCTGCTGCCTTGGCGA 585

Qy          183 TrpLeuAlaValLeuGlyCysValProProSerLeuMetLeuLeuLeuMetCysPheMet 202
                ||||||||||||   |||      |||      :::|||:::||||||:::   ||||||
Db          586 TGGCTTGCCGTGGCCGGGGAGGGGCCTGTTCTGGTCATGATCCTGCTGCTTAGCTTCATG 645

Qy          203 ProGluThrProArgPheLeuLeuThrGlnHisArgArgGlnGluAlaMetAlaAlaLeu 222
                |||   :::|||||||||||||||::::::   |||   :::||||||:::   ||||||
Db          646 CCCAACTCGCCTCGCTTCCTACTGTCTAAGAGCCGGGATGAGGAAGCACTGCAGGCGCTG 705

Qy          223 ArgPheLeuTrp---------GlySerGluGlnGlyTrpGluAspProProIleGly--- 238
                      :::|||            ||||||      ||||||         |||
Db          706 ------ATCTGGCTGCGAGCGGACTCTGAGGTCCACTGGGAGTTTGAGCAGATCCAGGAC 759

Qy          239 ------AlaGluGlnSerPheHisLeuAlaLeuLeuArg------GlnProGlyIleTyr 254
                            ||||||      ::::::               :::|||   :::|||
```

```
Db        760 AACGTGCGGAGACAGAGTAGCCGAGTGTCGTGGGCGGAGGCCTGGGAGCCCCGTGTGTAC 819

Qy        255 LysProPheIleIleGlyValSerLeuMetAlaPheGlnGlnLeuSerGlyValAsnAla 274
              :::|||   :::|||   |||   :::         |||||||||:::|||:::
Db        820 CGGCCTATTCTCATTACAGTGCTGATGCGTTTTCTGCAGCAGCTGACAGGCATCACTCCC 879

Qy        275 ValMetPheTyrAlaGluThrIlePheGluGluAlaLys---PheLysAspSerSerLeu 293
              ::::::   |||   :::|||||||||:::                     |||
Db        880 ATCCTCGTGTACCTACAGACCATCTTCGACAGCACGTCCGTGGTGCTGCCCTCCCAGCAG 939

Qy        294 AlaSerValValValGlyValIleGlnValLeuPheThrAlaValAlaAlaLeuIleMet 313
                 :::   :::||||||   :::::::::|||         :::||||||:::   |||
Db        940 GATGCAGCTATAGTTGGTGCTGTGAGGCTCCTGTCTGTGCTGATTGCCGCTGTCACCATG 999

Qy        314 AspArgAlaGlyArgArgLeuLeuLeuValLeuSerGlyValValMetValPheSerThr 333
              |||   |||||||||::::::||||||   :::|||      :::|||      :::
Db       1000 GACCTGGCCGGCCGAAAAGTCCTGCTCTATGTGTCAGCATCCATTATGTTTGTCGCCAAC 1059

Qy        334 SerAlaPheGlyAlaTyrPheLysLeuThrGlnGly------GlyProGlyAsnSerSer 351
                       |||   |||   :::|||                  |||   ::::::
Db       1060 CTGACGCTGGGGCTGTATGTCCAGCTTGTGCCAAGGACTCTGACCCCCAACAGCACTGTG 1119

Qy        352 HisValAlaIleSerAlaProValSerAlaGlnProValAspAlaSerValGlyLeuAla 371
                 :::                              |||      |||         |||
Db       1120 GAGATCGTGACGCTTGGGGGCACGGAGCAGCCCCCAGCCGCGGCCTTCAACTATCTCACC 1179

Qy        372 TrpLeuAlaValGlySerMetCysLeuPheIleAlaGlyPheAlaValGlyTrpGlyPro 391
                 :::   :::   :::      |||||||||   |||:::|||:::||||||||||||
Db       1180 CTGATACCCCTGCTGGCCACCATGCTCTTCATTATGGGCTATGCCATGGGCTGGGGGCCC 1239

Qy        392 IleProTrpLeuLeuMetSerGluIlePheProLeuHisValLysGlyValAlaThrGly 411
              |||   ||||||||||||||||||:::   ||||||         :::|||||||||:::|||
Db       1240 ATCACCTGGCTCCTCATGTCTGAGGTTCTGCCCCTGCGTGCCCGTGGTGTGGCCTCAGGG 1299

Qy        412 IleCysValLeuThrAsnTrpLeuMetAlaPheLeuValThrLysGluPheSerSerLeu 431
              :::|||||||||   :::||||||   ||||||::::::||||||   |||   :::
Db       1300 CTCTGCGTGCTGGTCAGCTGGCTCACAGCCTTCGTCCTCACTAAGTACTTCCT-GCTGGC 1358

Qy        432 MetGluValLeuArgProTyrGlyAlaPheTrpLeuAla-SerAlaPheCysIlePheSe 451
                 |||   |||||||||   |||||||||   |||    ||||||   |||:::   ||
Db       1359 AGTGAATGCCTTCGGCCTCCAGGTGCCTTTCTTCTTCTTCTCGGCCATCTGCCTGCTCAG 1418

Qy        451 rValLeuPheThrLeuPheCysValProGluThrLysGlyLysThrLeuGluGlnIleTh 471
              |:::|||||||||      |||||||||||||||:::|||::::::||||||||||||
Db       1419 CCTGCTCTTCACAGGCTGCTGTGTGCCTGAGACCAGGGGCCGCTCGCTGGAGCAGATCGA 1478

Qy        471 rAlaHisPheGluGlyArg 477
               |||   |||      |||
Db       1479 GGCCTTCTTCCACACTCGC 1497

RESULT 12
ADA84076
ID   ADA84076 standard; DNA; 2356 BP.
XX
AC   ADA84076;
XX
DT   20-NOV-2003  (first entry)
XX
DE   Human SLC2A6 gene.
XX
KW   human; marker; expressed sequence tag; EST; arabidopsis; tumour;
KW   stress-induced phenotype; hyperosmotic stress; colon cancer; immunogen;
KW   vaccine; ds; gene.
XX
OS   Homo sapiens.
XX
PN   WO2002103028-A2.
XX
PD   27-DEC-2002.
XX
PF   30-MAY-2002; 2002WO-IB004189.
XX
PR   30-MAY-2001; 2001US-0293999P.
```

```
PR   22-OCT-2001; 2001US-0330457P.
PR   19-FEB-2002; 2002US-0357144P.
XX
PA   (BIOM-) BIOMEDICAL CENT.
XX
PI   Baranova AV,  Yankovsky NK,  Kozlov AP,  Lobashev AV,  Krukovskaya LL;
XX
DR   WPI; 2003-175241/17.
DR   P-PSDB; ADA84077.
XX
PT   Determining if a nucleic acid is a marker for a phenotype/cell type of
PT   interest, by global comparison of expressed sequence tags known to be
PT   expressed in the phenotype/cell type with all ESTs expressed in normal
PT   tissue.
XX
PS   Claim 23; Page 457-458; 516pp; English.
XX
CC   The invention relates to a novel method for determining if a nucleic acid
CC   is a marker for a predetermined phenotype/cell type of interest from a
CC   biological species. The method comprises performing a global comparison
CC   of a group of expressed sequence tags (ESTs) known to be expressed in the
CC   phenotype/cell type of interest with all ESTs expressed in normal tissue
CC   in order to identify ESTs that are preferentially expressed in the
CC   phenotype/cell of interest. A method of the invention is useful for
CC   determining whether a nucleic acid is a marker for a predetermined
CC   phenotype or cell type of interest from a biological species, preferably
CC   Arabidopsis or human. The cell type of interest is an abnormal cell such
CC   as a tumour cell, and the predetermined phenotype is a stress-induced
CC   phenotype such as hyperosmotic stress or high salt conditions. A method
CC   of the invention is also useful for determining the progression of colon
CC   cancer in a human, for detecting a tumour cell, and for regulating or
CC   preventing the growth of a tumour cell. An antibody of the invention is
CC   useful for detecting the absence or presence of peptides encoded by
CC   tumour-associated markers. A polypeptide of the invention is useful as an
CC   immunogen for vaccinating an animal. The present sequence encodes a
CC   tumour-associated antigen of the invention.
XX
SQ   Sequence 2356 BP; 409 A; 786 C; 721 G; 440 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              1.18e-73      Length:         2356
Score:                  958.50        Matches:        227
Percent Similarity:     59.22%        Conservative:   78
Best Local Similarity:  44.08%        Mismatches:     164
Query Match:            39.01%        Indels:         47
DB:                     7             Gaps:           12

US-09-886-954A-1 (1-477) x ADA84076 (1-2356)

Qy          3 ProGlu---AspProGluGluThrGlnProLeuLeuGly--------------------- 14
              ||||||   ||||||         :::||||||||||||
Db         12 CCCGAGAGAGACCCGGCCATGCAGGAGCCGCTGCTGGGAGCCGAGGGCCCGGACTACGAC 71

Qy         15 ---------------ProProGlyGlySerAlaProArgGlyArg--------------- 24
                             ||||||               |||   |||
Db         72 ACCTTCCCCGAGAAGCCGCCCCCGTCGCCAGGGGACAGGGCGCGGGTCGGGACCCTGCAG 131

Qy         25 ------ArgValPheLeuAlaAlaPheAlaAlaAlaLeuGlyProLeuSerPheGlyPhe 42
                    |||||||||||||||   |||||||||   ||||||      |||||||||:::
Db        132 AACAAAAGGGTGTTCCTGGCCACCTTCGCCGCAGTGCTCGGCAATTTCAGCTTTGGGTAT 191

Qy         43 AlaLeuGlyTyrSerSerProAlaIleProSerLeuGlnArgAlaAlaProProAlaPro 62
              ||||||   |||:::||||||   ||||||:::|||:::|||:::      |||
Db        192 GCCCTGGTCTACACATCCCCTGTCATCCCAGCCCTGGAGCGCTCCTTGGATCCTGACCTG 251

Qy         63 ArgLeuAspAspAlaAlaAlaSerTrpPheGlyAlaValValThrLeuGlyAlaAlaAla 82
                 |||      :::   |||||||||||||||:::|||   ||||||||||||||||||
Db        252 CATCTGACCAAATCCCAGGCATCCTGGTTTGGGTCCGTGTTCACCCTGGGAGCAGCGGCC 311

Qy         83 GlyGlyValLeuGlyGlyTrpLeuValAspArgAlaGlyArgLysLeuSerLeuLeuLeu 102
              ||||||:::            |||   |||      |||||||||||||||::::::
Db        312 GGAGGCCTGAGTGCCATGATCCTCAACGACCTCCTGGGCCGGAAGCTGAGCATCATGTTC 371
```

```
Qy         103 CysSerValProPheValAlaGlyPheAlaValIleThrAlaAlaGlnAspValTrpMet 122
                  :::||||||      ||||||:::|||::::::      |||      :::||||||
Db         372 TCAGCTGTGCCGTCGGCGGCCGGCTATGCGCTCATGGCGGGTGCGCACGGCCTCTGGATG 431

Qy         123 LeuLeuGlyGlyArgLeuLeuThrGlyLeuAlaCysGlyValAlaSerLeuValAlaPro 142
               ||||||   ||||||   |||||||||   |||   |||:::   :::         |||
Db         432 CTGCTGCTCGGAAGGACGCTGACGGGCTTCGCCGGGGGGCTCACAGCTGCCTGCATCCCG 491

Qy         143 ValTyrIleSerGluIleAlaTyrProAlaValArgGlyLeuLeuGlySerCysValGln 162
               ||||||:::||||||||||||   |||   |||||||||   ||||||:::      |||
Db         492 GTGTACGTGTCTGAGATTGCTCCCCCAGGCGTTCGTGGGGCTCTGGGGGCCACACCCCAG 551

Qy         163 LeuMetValValValGlyIleLeuLeuAlaTyrLeuAlaGlyTrpValLeuGluTrpArg 182
               ||||||   |||   |||   |||      |||      |||   :::|||   ||||||
Db         552 CTCATGGCAGTGTTCGGATCCCTGTCCCTCTACGCCCTTGGCCTCCTGCTGCCGTGGCGC 611

Qy         183 TrpLeuAlaValLeuGlyCysValProProSerLeuMetLeuLeuLeuMetCysPheMet 202
               ||||||||||||   |||      |||      :::|||:::||||||:::   ||||||
Db         612 TGGCTGGCTGTGGCCGGGGAGGCGCCTGTGCTCATCATGATCCTGCTGCTCAGCTTCATG 671

Qy         203 ProGluThrProArgPheLeuLeuThrGlnHisArgArgGlnGluAlaMetAlaAlaLeu 222
               |||   :::|||||||||||||||::::::   |||   :::||||||:::   ||||||
Db         672 CCCAACTCGCCGCGCTTCCTGCTCTCTCGGGGCAGGGACGAAGAGGCCCTGCGGGCGCTG 731

Qy         223 ArgPheLeuTrpGlySerGluGlnGly------TrpGluAspProProIleGly------ 238
                  :::|||   |||::::::            ||||||         |||
Db         732 GCCTGGCTGCGTGGGACGGACGTCGATGTCCACTGGGAGTTCGAGCAGATCCAGGACAAC 791

Qy         239 ---AlaGluGlnSerPheHisLeu------AlaLeuLeuArgGlnProGlyIleTyrLys 255
                        ||||||      :::      |||      |||   |||   :::   :::
Db         792 GTCCGGAGACAGAGCAGCCGAGTATCGTGGGCTGAGGCACGGGCCCCCACGTGTGCCGG 851

Qy         256 ProPheIleIleGlyValSerLeuMetAlaPheGlnGlnLeuSerGlyValAsnAlaVal 275
               |||      :::   :::   :::         |||||||||:::|||:::      :::
Db         852 CCCATCACCGTGGCCTTGCTGATGCGCCTCCTGCAGCAGCTGACGGGCATCACGCCCATC 911

Qy         276 MetPheTyrAlaGluThrIlePheGluGluAlaLysPhe------------LysAspSer 291
               :::   |||   ::::::||||||:::                        ||||||
Db         912 CTGGTCTACCTGCAGTCCATCTTCGACAGCACCGCTGTCCTGCTGCCCCCCAAGGACGAC 971

Qy         292 SerLeuAlaSerValValValGlyValIleGlnValLeuPheThrAlaValAlaAlaLeu 311
                        :::   :::||||||   :::::::::|||         :::|||||||||
Db         972 ---------GCAGCCATCGTTGGGGCCGTGCGGCTCCTGTCCGTGCTGATCGCCGCCCTC 1022

Qy         312 IleMetAspArgAlaGlyArgArgLeuLeuLeuValLeuSerGlyValValMetValPhe 331
                  ||||||   |||||||||::::::||||||   :::|||      :::|||
Db        1023 ACCATGGACCTCGCAGGCCGCAAGGTGCTGCTCTTCGTCTCAGCGGCCATCATGTTTGCT 1082

Qy         332 SerThrSerAlaPheGlyAlaTyrPheLysLeuThrGlnGlyGlyProGlyAsnSerSer 351
               :::            |||   |||                  ||||||         |||
Db        1083 GCCAACCTGACTCTGGGGCTGTACATCCACTTT---------GGCCCCAGGCCTCTGAGC 1133

Qy         352 HisValAlaIleSerAlaProValSer------------------AlaGlnProValAsp 365
                     :::   :::         |||                  |||||||||:::
Db        1134 CCCAACAGCACTGCGGGCCTGGAAAGCGAGTCCTGGGGGACTTGGCGCAGCCCCTGGCA 1193

Qy         366 AlaSerValGly---LeuAlaTrpLeuAlaValGlySerMetCysLeuPheIleAlaGly 384
               |||      |||   |||      :::   :::   :::      |||||||||   |||
Db        1194 GCACCCGCTGGCTACCTCACCCTGGTGCCCCTGCTGGCCACCATGCTCTTCATCATGGGC 1253

Qy         385 PheAlaValGlyTrpGlyProIleProTrpLeuLeuMetSerGluIlePheProLeuHis 404
               :::|||||||||||||||||||||   ||||||||||||||||||:::   ||||||
Db        1254 TACGCCGTGGGCTGGGGTCCCATCACCTGGCTGCTCATGTCTGAGGTCCTGCCCCTGCGT 1313

Qy         405 ValLysGlyValAlaThrGlyIleCysValLeuThrAsnTrpLeuMetAlaPheLeuVal 424
                  :::|||||||||:::|||:::|||||||||   :::||||||   ||||||::::::
Db        1314 GCCCGTGGCGTGGCCTCAGGGCTCTGCGTGCTGGCCAGCTGGCTCACCGCCTTCGTCCTC 1373

Qy         425 ThrLysGluPheSerSerLeuMetGluValLeuArgProTyrGlyAlaPheTrpLeuAla 444
               ||||||   |||      :::   |||   |||||||||   |||||||||   |||
Db        1374 ACCAAGTCCTTCCTGCCAGTGGT-GAGCACCTTCGGCCTCCAGGTGCCTTTCTTCTTCTT 1432

Qy         445 -SerAlaPheCysIlePheSerValLeuPheThrLeuPheCysValProGluThrLysGl 464
                :::|||   |||:::   |||::::::||||||      ||||||||||||||||||||
Db        1433 CGCGGCCATCTGCTTGGTGAGCCTGGTGTTCACAGGCTGCTGTGTGCCCGAGACCAAGGG 1492

Qy         464 yLysThrLeuGluGlnIleThrAlaHisPheGlu---GlyArg 477
               |::::::||||||||||||   :::   |||      ||||||
Db        1493 ACGGTCCCTGGAGCAGATCGAGTCCTTCTTCCGCACGGGGAGA 1535

RESULT 13
ABZ24791
ID   ABZ24791 standard; cDNA; 2356 BP.
XX
AC   ABZ24791;
XX
DT   07-APR-2003  (first entry)
XX
DE   Human solute carrier type 2A nucleic acid 15489280.
XX
KW   Human; solute carrier type 2A; SLC2A; glucose transporter; p53; cancer;
KW   cytostatic; gene therapy; gene; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             30. .1553
FT                   /*tag=  a
FT                   /product= "Human SCL2A"
XX
PN   WO200298467-A1.
XX
PD   12-DEC-2002.
XX
PF   03-JUN-2002; 2002WO-US017419.
XX
PR   05-JUN-2001; 2001US-0296076P.
PR   10-OCT-2001; 2001US-0328605P.
PR   15-FEB-2002; 2002US-0357253P.
XX
PA   (EXEL-) EXELIXIS INC.
XX
PI   Friedman L,  Plowman GD,  Belvin M,  Francis-Lang H,  Li D,  Funke RP;
XX
DR   WPI; 2003-201283/19.
XX
PT   Identifying candidate p53 pathway modulating agent as therapeutic target
PT   for disorders of defective p53 function e.g. cancer, by assaying purified
PT   solute carrier family 2 (SLC2A) polypeptide or nucleic acid with a test
PT   agent.
XX
PS   Disclosure; Page 41-42; 58pp; English.
XX
CC   The present sequence is that of human solute carrier type 2A (SLC2A)
CC   nucleic acids 15489280. Genetic screens were designed to identify
CC   modifiers of the p53 pathway in Drosophila in which p53 was
CC   overexpressed. Human orthologues (polynucleotides and polypeptides) of
CC   one such modifier were then identified, including nucleic acid 15489280.
CC   SLC2As are glucose transporter proteins with sugar transporter domains.
CC   SLC2A nucleic acids and polypeptides are attractive drug targets for the
CC   treatment of pathologies associated with a defective p53 signalling
CC   pathway, such as cancer. The invention provides in vitro and in vivo
CC   methods of assessing SLC2A function. Modulation of an SLC2A or its
CC   binding partners is useful for understanding the association of the p53
CC   pathway and its members in normal and disease conditions and for
CC   developing diagnostic and therapeutic modalities of p53-related
CC   pathologies. SLC2A-modulating agents that act by inhibiting or enhancing
CC   SLC2A expression, directly or indirectly, e.g. by affecting an SLC2A
CC   function such as transport or binding activity, can be identified using
CC   methods provided. Modulators include small molecules, nucleic acids,
CC   antibodies, antisense oligonucleotides and phosphothioate morpholino
CC   oligomers (claimed)
XX
SQ   Sequence 2356 BP; 409 A; 786 C; 721 G; 440 T; 0 U; 0 Other;

Alignment Scores:
```

```
Pred. No.:                1.18e-73     Length:          2356
Score:                    958.50       Matches:         227
Percent Similarity:       59.22%       Conservative:    78
Best Local Similarity:    44.08%       Mismatches:      164
Query Match:              39.01%       Indels:          47
DB:                       7            Gaps:            12

US-09-886-954A-1 (1-477) x ABZ24791 (1-2356)

Qy            3 ProGlu---AspProGluGluThrGlnProLeuLeuGly--------------------- 14
                ||||||   ||||||         :::||||||||||||
Db           12 CCCGAGAGAGACCCGGCCATGCAGGAGCCGCTGCTGGGAGCCGAGGGCCCGGACTACGAC 71

Qy           15 ---------------ProProGlyGlySerAlaProArgGlyArg--------------- 24
                               ||||||               |||   |||
Db           72 ACCTTCCCCGAGAAGCCGCCCCCGTCGCCAGGGGACAGGGCGCGGGTCGGGACCCTGCAG 131

Qy           25 ------ArgValPheLeuAlaAlaPheAlaAlaAlaLeuGlyProLeuSerPheGlyPhe 42
                      |||||||||||||||   |||||||||   ||||||      |||||||||:::
Db          132 AACAAAAGGGTGTTCCTGGCCACCTTCGCCGCAGTGCTCGGCAATTTCAGCTTTGGGTAT 191

Qy           43 AlaLeuGlyTyrSerSerProAlaIleProSerLeuGlnArgAlaAlaProProAlaPro 62
                ||||||   |||:::||||||   ||||||:::|||:::|||:::      |||
Db          192 GCCCTGGTCTACACATCCCCTGTCATCCCAGCCCTGGAGCGCTCCTTGGATCCTGACCTG 251

Qy           63 ArgLeuAspAspAlaAlaAlaSerTrpPheGlyAlaValValThrLeuGlyAlaAlaAla 82
                   |||      :::   |||||||||||||||:::|||   ||||||||||||||||||
Db          252 CATCTGACCAAATCCCAGGCATCCTGGTTTGGGTCCGTGTTCACCCTGGGAGCAGCGGCC 311

Qy           83 GlyGlyValLeuGlyGlyTrpLeuValAspArgAlaGlyArgLysLeuSerLeuLeuLeu 102
                ||||||:::            |||   |||      |||||||||||||||::::::
Db          312 GGAGGCCTGAGTGCCATGATCCTCAACGACCTCCTGGGCCGGAAGCTGAGCATCATGTTC 371

Qy          103 CysSerValProPheValAlaGlyPheAlaValIleThrAlaAlaGlnAspValTrpMet 122
                   :::||||||      ||||||:::|||::::::      |||      :::||||||
Db          372 TCAGCTGTGCCGTCGGCGGCCGGCTATGCGCTCATGGCGGGTGCGCACGGCCTCTGGATG 431

Qy          123 LeuLeuGlyGlyArgLeuLeuThrGlyLeuAlaCysGlyValAlaSerLeuValAlaPro 142
                ||||||   ||||||   |||||||||   |||   |||:::   :::         |||
Db          432 CTGCTGCTCGGAAGGACGCTGACGGGCTTCGCCGGGGGGCTCACAGCTGCCTGCATCCCG 491

Qy          143 ValTyrIleSerGluIleAlaTyrProAlaValArgGlyLeuLeuGlySerCysValGln 162
                ||||||:::||||||||||||   |||   |||||||||   ||||||:::      |||
Db          492 GTGTACGTGTCTGAGATTGCTCCCCCAGGCGTTCGTGGGGCTCTGGGGGCCACACCCCAG 551

Qy          163 LeuMetValValValGlyIleLeuLeuAlaTyrLeuAlaGlyTrpValLeuGluTrpArg 182
                ||||||   |||   |||   |||      |||      |||   :::|||   ||||||
Db          552 CTCATGGCAGTGTTCGGATCCCTGTCCCTCTACGCCCTTGGCCTCCTGCTGCCGTGGCGC 611

Qy          183 TrpLeuAlaValLeuGlyCysValProProSerLeuMetLeuLeuLeuMetCysPheMet 202
                ||||||||||||   |||      |||      :::|||:::||||||:::   ||||||
Db          612 TGGCTGGCTGTGGCCGGGGAGGCGCCTGTGCTCATCATGATCCTGCTGCTCAGCTTCATG 671

Qy          203 ProGluThrProArgPheLeuLeuThrGlnHisArgArgGlnGluAlaMetAlaAlaLeu 222
                |||   :::|||||||||||||||::::::   |||   :::||||||:::   ||||||
Db          672 CCCAACTCGCCGCGCTTCCTGCTCTCTCGGGGCAGGGACGAAGAGGCCCTGCGGGCGCTG 731

Qy          223 ArgPheLeuTrpGlySerGluGlnGly------TrpGluAspProProIleGly------ 238
                   :::|||   |||::::::            ||||||         |||
Db          732 GCCTGGCTGCGTGGACGGACGTCGATGTCCACTGGGAGTTCGAGCAGATCCAGGACAAC 791

Qy          239 ---AlaGluGlnSerPheHisLeu------AlaLeuLeuArgGlnProGlyIleTyrLys 255
                         ||||||      :::      |||      |||   |||   :::   :::
Db          792 GTCCGGAGACAGAGCAGCCGAGTATCGTGGGCTGAGGCACGGGCCCCCCACGTGTGCCGG 851

Qy          256 ProPheIleIleGlyValSerLeuMetAlaPheGlnGlnLeuSerGlyValAsnAlaVal 275
                |||      :::   :::   :::         |||||||||:::|||:::      :::
Db          852 CCCATCACCGTGGCCTTGCTGATGCGCCTCCTGCAGCAGCTGACGGGCATCACGCCCATC 911

Qy          276 MetPheTyrAlaGluThrIlePheGluGluAlaLysPhe------------LysAspSer 291
                :::   |||   ::::::||||||:::                        ||||||
Db          912 CTGGTCTACCTGCAGTCCATCTTCGACAGCACCGCTGTCCTGCTGCCCCCCAAGGACGAC 971

Qy          292 SerLeuAlaSerValValValGlyValIleGlnValLeuPheThrAlaValAlaAlaLeu 311
                         :::   :::||||||   :::::::::|||         :::|||||||||
Db          972 ---------GCAGCCATCGTTGGGGCCGTGCGGCTCCTGTCCGTGCTGATCGCCGCCCTC 1022

Qy          312 IleMetAspArgAlaGlyArgArgLeuLeuLeuValLeuSerGlyValValMetValPhe 331
                   ||||||   |||||||||::::::||||||   :::|||      :::|||
Db         1023 ACCATGGACCTCGCAGGCCGCAAGGTGCTGCTCTTCGTCTCAGCGGCCATCATGTTTGCT 1082

Qy          332 SerThrSerAlaPheGlyAlaTyrPheLysLeuThrGlnGlyGlyProGlyAsnSerSer 351
                :::            |||   |||                  ||||||         |||
Db         1083 GCCAACCTGACTCTGGGGCTGTACATCCACTTT---------GGCCCCAGGCCTCTGAGC 1133

Qy          352 HisValAlaIleSerAlaProValSer------------------AlaGlnProValAsp 365
                      :::   :::         |||                  |||||||||:::
Db         1134 CCCAACAGCACTGCGGGCCTGGAAAGCGAGTCCTGGGGGGACTTGGCGCAGCCCCTGGCA 1193

Qy          366 AlaSerValGly---LeuAlaTrpLeuAlaValGlySerMetCysLeuPheIleAlaGly 384
                |||      |||   |||      :::   :::   :::      |||||||||   |||
Db         1194 GCACCCGCTGGCTACCTCACCCTGGTGCCCCTGCTGGCCACCATGCTCTTCATCATGGGC 1253

Qy          385 PheAlaValGlyTrpGlyProIleProTrpLeuLeuMetSerGluIlePheProLeuHis 404
                :::|||||||||||||||||||||   ||||||||||||||||||:::   ||||||
Db         1254 TACGCCGTGGGCTGGGGTCCCATCACCTGGCTGCTCATGTCTGAGGTCCTGCCCCTGCGT 1313

Qy          405 ValLysGlyValAlaThrGlyIleCysValLeuThrAsnTrpLeuMetAlaPheLeuVal 424
                   :::|||||||||:::|||:::|||||||||   :::||||||   ||||||::::::
Db         1314 GCCCGTGGCGTGGCCTCAGGGCTCTGCGTGCTGGCCAGCTGGCTCACCGCCTTCGTCCTC 1373

Qy          425 ThrLysGluPheSerSerLeuMetGluValLeuArgProTyrGlyAlaPheTrpLeuAla 444
                ||||||   |||      :::   |||   |||||||||   |||||||||   |||
Db         1374 ACCAAGTCCTTCCTGCCAGTGGT-GAGCACCTTCGGCCTCCAGGTGCCTTTCTTCTTCTT 1432

Qy          445 -SerAlaPheCysIlePheSerValLeuPheThrLeuPheCysValProGluThrLysGl 464
                 :::|||   |||:::   |||::::::||||||      ||||||||||||||||||||
Db         1433 CGCGGCCATCTGCTTGGTGAGCCTGGTGTTCACAGGCTGCTGTGTGCCCGAGACCAAGGG 1492

Qy          464 yLysThrLeuGluGlnIleThrAlaHisPheGlu---GlyArg 477
                |::::::||||||||||||   :::   |||      ||||||
Db         1493 ACGGTCCCTGGAGCAGATCGAGTCCTTCTTCCGCACGGGGAGA 1535

RESULT 14
AAK94338
ID   AAK94338 standard; cDNA; 2390 BP.
XX
AC   AAK94338;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human full-length cDNA, SEQ ID NO: 3033.
XX
KW   Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1130094-A2.
XX
PD   05-SEP-2001.
XX
PF   07-JUL-2000; 2000EP-00114089.
XX
PR   08-JUL-1999;    99JP-00194486.
PR   11-JAN-2000; 2000JP-00118774.
PR   02-MAY-2000; 2000JP-00183765.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Nishikawa T,  Isogai T,  Hayashi K,  Ishii S,  Kawai Y;
PI   Wakamatsu A,  Sugiyama T,  Nagai K,  Kojima S,  Otsuki T,  Koga H;
XX
DR   WPI; 2001-524255/58.
DR   P-PSDB; AAM93417.
XX
PT   830 Primers useful for synthesizing full length cDNA clones and their use
```

```
PT   in genetic manipulation.
XX
PS   Claim 8; SEQ ID NO 3033; 1380pp + Sequence Listing; English.
XX
CC   The invention relates to primers for synthesising full length cDNA
CC   clones. 830 cDNA molecules encoding a human protein have been isolated
CC   and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have
CC   been determined. Primers for synthesising the full length cDNA are useful
CC   for clarifying the function of the protein encoded by the cDNA. The full
CC   length clones were obtained by construction of full length enriched cDNA
CC   libraries that were synthesised by the oligo-capping method. The primers
CC   enable the production of the full length cDNA easily without any special
CC   methods. The present sequence is a full length human cDNA of the
CC   invention. Note: The sequence data for this patent did not form part of
CC   the printed specification, but was obtained in CD-ROM format directly
CC   from EPO
XX
SQ   Sequence 2390 BP; 398 A; 803 C; 739 G; 450 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              1.21e-73        Length:         2390
Score:                  958.50          Matches:        227
Percent Similarity:     59.22%          Conservative:   78
Best Local Similarity:  44.08%          Mismatches:     164
Query Match:            39.01%          Indels:         47
DB:                     4               Gaps:           12

US-09-886-954A-1 (1-477) x AAK94338 (1-2390)

Qy          3 ProGlu---AspProGluGluThrGlnProLeuLeuGly--------------------- 14
              ||||||   ||||||         :::||||||||||||
Db         61 CCCGAGAGAGACCCGGCCATGCAGGAGCCGCTGCTGGGAGCCGAGGGCCCGGACTACGAC 120

Qy         15 ---------------ProProGlyGlySerAlaProArgGlyArg--------------- 24
                             ||||||               |||   |||
Db        121 ACCTTCCCCGAGAAGCCGCCCCCGTCGCCAGGGGACAGGGCGCGGTCGGGACCCTGCAG 180

Qy         25 ------ArgValPheLeuAlaAlaPheAlaAlaAlaLeuGlyProLeuSerPheGlyPhe 42
                    |||||||||||||||   |||||||||   ||||||      |||||||||:::
Db        181 AACAAAAGGGTGTTCCTGGCCACCTTCGCCGCAGTGCTCGGCAATTTCAGCTTTGGGTAT 240

Qy         43 AlaLeuGlyTyrSerSerProAlaIleProSerLeuGlnArgAlaAlaProProAlaPro 62
              ||||||   |||:::||||||   ||||||:::|||:::|||:::      |||
Db        241 GCCCTGGTCTACACATCCCCTGTCATCCCAGCCCTGGAGCGCTCCTTGGATCCTGACCTG 300

Qy         63 ArgLeuAspAspAlaAlaAlaSerTrpPheGlyAlaValValThrLeuGlyAlaAlaAla 82
                 |||      :::   |||||||||||||||:::|||   ||||||||||||||||||
Db        301 CATCTGACCAAATCCCAGGCATCCTGGTTTGGGTCCGTGTTCACCCTGGGAGCAGCGGCC 360

Qy         83 GlyGlyValLeuGlyGlyTrpLeuValAspArgAlaGlyArgLysLeuSerLeuLeuLeu 102
              ||||||:::            |||   |||      |||||||||||||||::::::
Db        361 GGAGGCCTGAGTGCCATGATCCTCAACGACCTCCTGGGCCGGAAGCTGAGCATCATGTTC 420

Qy        103 CysSerValProPheValAlaGlyPheAlaValIleThrAlaAlaGlnAspValTrpMet 122
                 :::||||||      ||||||:::|||::::::      |||      :::||||||
Db        421 TCAGCTGTGCCGTCGGCGGCCGGCTATGCGCTCATGGCGGGTGCGCACGGCCTCTGGATG 480

Qy        123 LeuLeuGlyGlyArgLeuLeuThrGlyLeuAlaCysGlyValAlaSerLeuValAlaPro 142
              ||||||   ||||||   |||||||||   |||   |||:::   :::         |||
Db        481 CTGCTGCTCGGAAGGACGCTGACGGGCTTCGCCGGGGGCTCACAGCTGCCTGCATCCCG 540

Qy        143 ValTyrIleSerGluIleAlaTyrProAlaValArgGlyLeuLeuGlySerCysValGln 162
              ||||||:::||||||||||||   |||   |||||||||   ||||||:::      |||
Db        541 GTGTACGTGTCTGAGATTGCTCCCCCAGGCGTTCGTGGGGCTCTGGGGGCCACACCCCAG 600

Qy        163 LeuMetValValValGlyIleLeuLeuAlaTyrLeuAlaGlyTrpValLeuGluTrpArg 182
              ||||||   |||   |||   |||      |||      |||   :::|||   ||||||
Db        601 CTCATGGCAGTGTTCGGATCCCTGTCCCTCTACGCCCTTGGCCTCCTGCTGCCGTGGCGC 660

Qy        183 TrpLeuAlaValLeuGlyCysValProProSerLeuMetLeuLeuLeuMetCysPheMet 202
              ||||||||||||   |||      |||      :::|||:::||||||:::   ||||||
Db        661 TGGCTGGCTGTGGCCGGGGAGGCGCCTGTGCTCATCATGATCCTGCTGCTCAGCTTCATG 720

Qy        203 ProGluThrProArgPheLeuLeuThrGlnHisArgArgGlnGluAlaMetAlaAlaLeu 222
              |||   :::|||||||||||||||::::::   |||   :::||||||:::   ||||||
Db        721 CCCAACTCGCCGCGCTTCCTGCTCTCGGGGCAGGGACGAAGAGGCCCTGCGGGCGCTG 780

Qy        223 ArgPheLeuTrpGlySerGluGlnGly------TrpGluAspProProIleGly------ 238
                 :::|||   |||::::::            ||||||         |||
Db        781 GCCTGGCTGCGTGGGACGGACGTCGATGTCCACTGGGAGTTCGAGCAGATCCAGGACAAC 840

Qy        239 ---AlaGluGlnSerPheHisLeu------AlaLeuLeuArgGlnProGlyIleTyrLys 255
                       ||||||      :::      |||      |||   |||   :::   :::
Db        841 GTCCGGAGACAGAGCAGCCGAGTATCGTGGGCTGAGGCACGGGCCCCACACGTGTGCCGG 900

Qy        256 ProPheIleIleGlyValSerLeuMetAlaPheGlnGlnLeuSerGlyValAsnAlaVal 275
              |||      :::   :::   :::         |||||||||:::|||:::      :::
Db        901 CCCATCACCGTGGCCTTGCTGATGCGCCTCCTGCAGCAGCTGACGGGCATCACGCCCATC 960

Qy        276 MetPheTyrAlaGluThrIlePheGluGluAlaLysPhe------------LysAspSer 291
              :::   |||   ::::::||||||:::                        ||||||
Db        961 CTGGTCTACCTGCAGTCTATCTTCGACAGCACCGCTGTCCTGCTGCCCCCCAAGGACGAC 1020

Qy        292 SerLeuAlaSerValValValGlyValIleGlnValLeuPheThrAlaValAlaAlaLeu 311
                       :::   :::||||||   :::::::::|||         :::|||||||||
Db       1021 ---------GCAGCCATCGTTGGGGCCGTGCGGCTCCTGTCCGTGCTGATCGCCGCCCTC 1071

Qy        312 IleMetAspArgAlaGlyArgArgLeuLeuLeuValLeuSerGlyValValMetValPhe 331
                 ||||||   |||||||||::::::||||||   :::|||      :::|||
Db       1072 ACCATGGACCTCGCAGGCCGCAAGGTGCTGCTCTTCGTCTCAGCGGCCATCATGTTTGCT 1131

Qy        332 SerThrSerAlaPheGlyAlaTyrPheLysLeuThrGlnGlyGlyProGlyAsnSerSer 351
              :::            |||   |||                  ||||||         |||
Db       1132 GCCAACCTGACTCTGGGGCTGTACATCCACTTT---------GGCCCCAGGCCTCTGAGC 1182

Qy        352 HisValAlaIleSerAlaProValSer------------------AlaGlnProValAsp 365
                    :::   :::         |||                  |||||||||:::
Db       1183 CCCAACAGCACTGCGGGCCTGGAAAGCGAGTCCTGGGGGGACTTGGCGCAGCCCCTGGCA 1242

Qy        366 AlaSerValGly---LeuAlaTrpLeuAlaValGlySerMetCysLeuPheIleAlaGly 384
              |||      |||   |||      :::   :::   :::      |||||||||   |||
Db       1243 GCACCCGCTGGCTACCTCACCCTGGTGCCCCTGCTGGCCACCATGCTCTTCATCATGGGC 1302

Qy        385 PheAlaValGlyTrpGlyProIleProTrpLeuLeuMetSerGluIlePheProLeuHis 404
              :::|||||||||||||||||||||   ||||||||||||||||||:::   ||||||
Db       1303 TACGCCGTGGGCTGGGGTCCCATCACCTGGCTGCTCATGTCTGAGGTCCTGCCCCTGCGT 1362

Qy        405 ValLysGlyValAlaThrGlyIleCysValLeuThrAsnTrpLeuMetAlaPheLeuVal 424
                 :::|||||||||:::|||:::|||||||||   :::||||||   ||||||::::::
Db       1363 GCCCGTGGCGTGGCCTCAGGGCTCTGCGTGCTGGCCAGCTGGCTCACCGCCTTCGTCCTC 1422

Qy        425 ThrLysGluPheSerSerLeuMetGluValLeuArgProTyrGlyAlaPheTrpLeuAla 444
              ||||||   |||      :::   |||   |||||||||   |||||||||   |||
Db       1423 ACCAAGTCCTTCCTGCCAGTGGT-GAGCACCTTCGGCCTCCAGGTGCCTTTCTTCTTCTT 1481

Qy        445 -SerAlaPheCysIlePheSerValLeuPheThrLeuPheCysValProGluThrLysGl 464
               :::|||   |||:::   |||::::::||||||      ||||||||||||||||||||
Db       1482 CGCGGCCATCTGCTTGGTGAGCCTGGTGTTCACAGGCTGCTGTGTGCCCGAGACCAAGGG 1541

Qy        464 yLysThrLeuGluGlnIleThrAlaHisPheGlu---GlyArg 477
              |::::::||||||||||||   :::   |||      ||||||
Db       1542 ACGGTCCCTGGAGCAGATCGAGTCCTTCTTCCGCACGGGGAGA 1584

RESULT 15
ABZ24790
ID   ABZ24790 standard; cDNA; 2487 BP.
XX
AC   ABZ24790;
XX
DT   07-APR-2003  (first entry)
XX
DE   Human solute carrier type 2A nucleic acid 12735203 and 9938031.
XX
KW   Human; solute carrier type 2A; SLC2A; glucose transporter; p53; cancer;
KW   cytostatic; gene therapy; gene; ss.
XX
```

```
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             38. .1561
FT                   /*tag=  a
FT                   /product= "Human SCL2A"
XX
PN   WO200298467-A1.
XX
PD   12-DEC-2002.
XX
PF   03-JUN-2002; 2002WO-US017419.
XX
PR   05-JUN-2001; 2001US-0296076P.
PR   10-OCT-2001; 2001US-0328605P.
PR   15-FEB-2002; 2002US-0357253P.
XX
PA   (EXEL-) EXELIXIS INC.
XX
PI   Friedman L,  Plowman GD,  Belvin M,  Francis-Lang H,  Li D,  Funke RP;
XX
DR   WPI; 2003-201283/19.
DR   P-PSDB; ABP58363.
XX
PT   Identifying candidate p53 pathway modulating agent as therapeutic target
PT   for disorders of defective p53 function e.g. cancer, by assaying purified
PT   solute carrier family 2 (SLC2A) polypeptide or nucleic acid with a test
PT   agent.
XX
PS   Disclosure; Page 38-39; 58pp; English.
XX
CC   The present sequence is that of human solute carrier type 2A (SLC2A)
CC   nucleic acids 12735203 and 9938031. In the present invention, genetic
CC   screens were designed to identify modifiers of the p53 pathway in
CC   Drosophila in which p53 was overexpressed. Human orthologues
CC   (polynucleotides and polypeptides) of one such modifier were then
CC   identified, including nucleic acids 12735203 and 9938031. SLC2As are
CC   glucose transporter proteins with sugar transporter domains. SLC2A
CC   nucleic acids and polypeptides are attractive drug targets for the
CC   treatment of pathologies associated with a defective p53 signalling
CC   pathway, such as cancer. The invention provides in vitro and in vivo
CC   methods of assessing SLC2A function. Modulation of an SLC2A or its
CC   binding partners is useful for understanding the association of the p53
CC   pathway and its members in normal and disease conditions and for
CC   developing diagnostic and therapeutic modalities of p53-related
CC   pathologies. SLC2A-modulating agents that act by inhibiting or enhancing
CC   SLC2A expression, directly or indirectly, e.g. by affecting an SLC2A
CC   function such as transport or binding activity, can be identified using
CC   methods provided. Modulators include small molecules, nucleic acids,
CC   antibodies, antisense oligonucleotides and phosphothioate morpholino
CC   oligomers (claimed)
XX
SQ   Sequence 2487 BP; 443 A; 826 C; 752 G; 466 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              1.27e-73        Length:         2487
Score:                  958.50          Matches:        227
Percent Similarity:     59.22%          Conservative:   78
Best Local Similarity:  44.08%          Mismatches:     164
Query Match:            39.01%          Indels:         47
DB:                     7               Gaps:           12

US-09-886-954A-1 (1-477) x ABZ24790 (1-2487)

Qy          3 ProGlu---AspProGluGluThrGlnProLeuLeuGly--------------------- 14
              ||||||   ||||||         :::||||||||||||
Db         20 CCCGAGAGAGACCCGGCCATGCAGGAGCCGCTGCTGGGAGCCGAGGGCCCGGACTACGAC 79

Qy         15 ---------------ProProGlyGlySerAlaProArgGlyArg--------------- 24
                             ||||||               |||   |||
Db         80 ACCTTCCCCGAGAAGCCGCCCCCGTCGCCAGGGGACAGGGCGCGGTCGGGACCCTGCAG 139

Qy         25 ------ArgValPheLeuAlaAlaPheAlaAlaAlaLeuGlyProLeuSerPheGlyPhe 42
                    |||||||||||||||   |||||||||   ||||||      |||||||||:::
Db        140 AACAAAAGGGTGTTCCTGGCCACCTTCGCCGCAGTGCTCGGCAATTTCAGCTTTGGGTAT 199

Qy         43 AlaLeuGlyTyrSerSerProAlaIleProSerLeuGlnArgAlaAlaProProAlaPro 62
              ||||||   |||:::||||||   ||||||:::|||:::|||:::      |||
Db        200 GCCCTGGTCTACACATCCCCTGTCATCCCAGCCCTGGAGCGCTCCTTGGATCCTGACCTG 259

Qy         63 ArgLeuAspAspAlaAlaAlaSerTrpPheGlyAlaValValThrLeuGlyAlaAlaAla 82
                 |||      :::   |||||||||||||||:::|||   ||||||||||||||||||
Db        260 CATCTGACCAAATCCCAGGCATCCTGGTTTGGGTCCGTGTTCACCCTGGGAGCAGCGGCC 319

Qy         83 GlyGlyValLeuGlyGlyTrpLeuValAspArgAlaGlyArgLysLeuSerLeuLeuLeu 102
              ||||||:::            |||   |||      |||||||||||||||::::::
Db        320 GGAGGCCTGAGTGCCATGATCCTCAACGACCTCCTGGGCCGGAAGCTGAGCATCATGTTC 379

Qy        103 CysSerValProPheValAlaGlyPheAlaValIleThrAlaAlaGlnAspValTrpMet 122
                 :::||||||      ||||||:::|||::::::      |||      :::||||||
Db        380 TCAGCTGTGCCGTCGGCGGCCGGCTATGCGCTCATGGCGGGTGCGCACGGCCTCTGGATG 439

Qy        123 LeuLeuGlyGlyArgLeuLeuThrGlyLeuAlaCysGlyValAlaSerLeuValAlaPro 142
              ||||||   ||||||   |||||||||   |||   |||:::   :::         |||
Db        440 CTGCTGCTCGGAAGGACGCTGACGGGCTTCGCCGGGGGGCTCACAGCTGCCTGCATCCCG 499

Qy        143 ValTyrIleSerGluIleAlaTyrProAlaValArgGlyLeuLeuGlySerCysValGln 162
              ||||||:::||||||||||||   |||   |||||||||   ||||||:::      |||
Db        500 GTGTACGTGTCTGAGATTGCTCCCCCAGGCGTTCGTGGGGCTCTGGGGGCCACACCCCAG 559

Qy        163 LeuMetValValValGlyIleLeuLeuAlaTyrLeuAlaGlyTrpValLeuGluTrpArg 182
              ||||||   |||   |||   |||      |||      |||   :::|||   ||||||
Db        560 CTCATGGCAGTGTTCGGATCCCTGTCCCTCTACGCCCTTGGCCTCCTGCTGCCGTGGCGC 619

Qy        183 TrpLeuAlaValLeuGlyCysValProProSerLeuMetLeuLeuLeuMetCysPheMet 202
              ||||||||||||   |||      |||      :::|||:::||||||:::   ||||||
Db        620 TGGCTGGCTGTGGCCGGGGAGGCGCCTGTGCTCATCATGATCCTGCTGCTCAGCTTCATG 679

Qy        203 ProGluThrProArgPheLeuLeuThrGlnHisArgArgGlnGluAlaMetAlaAlaLeu 222
              |||   :::|||||||||||||||::::::   |||   :::||||||:::   ||||||
Db        680 CCCAACTCGCCGCGCTTCCTGCTCTCTCGGGGCAGGGACGAAGAGGCCCTGCGGGCGCTG 739

Qy        223 ArgPheLeuTrpGlySerGluGlnGly------TrpGluAspProProIleGly------ 238
                 :::|||   |||::::::            ||||||         |||
Db        740 GCCTGGCTGCGTGGACGGACGTCGATGTCCACTGGGAGTTCGAGCAGATCCAGGACAAC 799

Qy        239 ---AlaGluGlnSerPheHisLeu------AlaLeuLeuArgGlnProGlyIleTyrLys 255
                       ||||||      :::      |||      |||   |||   :::   :::
Db        800 GTCCGGAGACAGAGCAGCCGAGTATCGTGGGCTGAGGCACGGGCCCCACACGTGTGCCGG 859

Qy        256 ProPheIleIleGlyValSerLeuMetAlaPheGlnGlnLeuSerGlyValAsnAlaVal 275
              |||      :::   :::   :::         |||||||||:::|||:::      :::
Db        860 CCCATCACCGTGGCCTTGCTGATGCGCCTCCTGCAGCAGCTGACGGGCATCACGCCCATC 919

Qy        276 MetPheTyrAlaGluThrIlePheGluGluAlaLysPhe------------LysAspSer 291
              :::   |||   ::::::||||||:::                        ||||||
Db        920 CTGGTCTACCTGCAGTCCATCTTCGACAGCACCGCTGTCCTGCTGCCCCCCAAGGACGAC 979

Qy        292 SerLeuAlaSerValValValGlyValIleGlnValLeuPheThrAlaValAlaAlaLeu 311
                       :::   :::||||||   :::::::::|||         :::|||||||||
Db        980 ---------GCAGCCATCGTTGGGGCCGTGCGGCTCCTGTCCGTGCTGATCGCCGCCCTC 1030

Qy        312 IleMetAspArgAlaGlyArgArgLeuLeuLeuValLeuSerGlyValValMetValPhe 331
                 ||||||   |||||||||::::::||||||   :::|||      :::|||
Db       1031 ACCATGGACCTCGCAGGCCGCAAGGTGCTGCTCTTCGTCTCAGCGGCCATCATGTTTGCT 1090

Qy        332 SerThrSerAlaPheGlyAlaTyrPheLysLeuThrGlnGlyGlyProGlyAsnSerSer 351
              :::            |||   |||                  ||||||         |||
Db       1091 GCCAACCTGACTCTGGGGCTGTACATCCACTTT---------GGCCCCAGGCCTCTGAGC 1141

Qy        352 HisValAlaIleSerAlaProValSer------------------AlaGlnProValAsp 365
                    :::   :::         |||                  |||||||||:::
Db       1142 CCCAACAGCACTGCGGGCCTGGAAAGCGAGTCCTGGGGGGACTTGGCGCAGCCCCTGGCA 1201

Qy        366 AlaSerValGly---LeuAlaTrpLeuAlaValGlySerMetCysLeuPheIleAlaGly 384
              |||      |||   |||      :::   :::   :::      |||||||||   |||
```

```
Db          1202 GCACCCGCTGGCTACCTCACCCTGGTGCCCCTGCTGGCCACCATGCTCTTCATCATGGGC 1261

Qy           385 PheAlaValGlyTrpGlyProIleProTrpLeuLeuMetSerGluIlePheProLeuHis 404
                 :::|||||||||||||||||||||   ||||||||||||||||||:::   ||||||
Db          1262 TACGCCGTGGGCTGGGGTCCCATCACCTGGCTGCTCATGTCTGAGGTCCTGCCCCTGCGT 1321

Qy           405 ValLysGlyValAlaThrGlyIleCysValLeuThrAsnTrpLeuMetAlaPheLeuVal 424
                    :::|||||||||:::|||:::|||||||||   :::||||||   ||||||::::::
Db          1322 GCCCGTGGCGTGGCCTCAGGGCTCTGCGTGCTGGCCAGCTGGCTCACCGCCTTCGTCCTC 1381

Qy           425 ThrLysGluPheSerSerLeuMetGluValLeuArgProTyrGlyAlaPheTrpLeuAla 444
                 ||||||   |||      :::   |||   |||||||||   |||||||||   |||
Db          1382 ACCAAGTCCTTCCTGCCAGTGGT-GAGCACCTTCGGCCTCCAGGTGCCTTTCTTCTTCTT 1440

Qy           445 -SerAlaPheCysIlePheSerValLeuPheThrLeuPheCysValProGluThrLysGl 464
                  :::|||   |||:::   |||::::::||||||      ||||||||||||||||||||
Db          1441 CGCGGCCATCTGCTTGGTGAGCCTGGTGTTCACAGGCTGCTGTGTGCCCGAGACCAAGGG 1500

Qy           464 yLysThrLeuGluGlnIleThrAlaHisPheGlu---GlyArg 477
                 |::::::||||||||||||   :::   |||      ||||||
Db          1501 ACGGTCCCTGGAGCAGATCGAGTCCTTCTTCCGCATGGGGAGA 1543

Search completed: September 28, 2004, 02:12:01
Job time : 518 secs
```

```
                    GenCore version 5.1.6
           Copyright (c) 1993 - 2004  Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on:         September 28, 2004, 01:55:59 ; Search time 111 Seconds
                                           (without alignments)
                                           2384.789 Million cell updates/sec

Title:          US-09-886-954A-1
Perfect score:  2457
Sequence:       1 MTPEDPEETQPLLGPPGGSA..........CVPETKGKTLEQITAHFEGR 477

Scoring table:  BLOSUM62
                Xgapop 10.0 , Xgapext  0.5
                Ygapop 10.0 , Ygapext  0.5
                Fgapop  6.0 , Fgapext  7.0
                Delop   6.0 , Delext   7.0

Searched:       682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters:      1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US09886954/runat_27092004_170632_13412/app_query.fasta_1.647
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09886954_@CGN_1_1_105_@runat_27092004_170632_13412 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :      Issued_Patents_NA:*
                1:  /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
                2:  /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
                3:  /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
                4:  /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
                5:  /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
                6:  /cgn2_6/ptodata/2/ina/backfiles1.seq:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                   SUMMARIES
                 %
Result           Query
  No.    Score   Match Length DB  ID                        Description
-------------------------------------------------------------------------------
    1      605    24.6   1545  4  US-09-489-039A-4731        Sequence 4731, Ap
    2    578.5    23.5   1431  4  US-09-489-039A-4762        Sequence 4762, Ap
    3    570.5    23.2   3915  4  US-09-023-655-1104         Sequence 1104, Ap
    4    542.5    22.1   1752  4  US-09-679-686B-17          Sequence 17, Appl
    5    532.5    21.7   1675  4  US-09-679-686B-1           Sequence 1, Appli
    6      518    21.1   1960  4  US-09-679-686B-15          Sequence 15, Appl
    7    514.5    20.9   2856  4  US-09-643-597-135          Sequence 135, App
    8    514.5    20.9   2856  4  US-09-480-884A-135         Sequence 135, App
    9    514.5    20.9   2856  4  US-09-542-615A-135         Sequence 135, App
   10    514.5    20.9   2856  4  US-09-606-421B-135         Sequence 135, App
   11    514.5    20.9   2856  4  US-09-221-107-135          Sequence 135, App
   12    507.5    20.7   1506  4  US-09-489-039A-4560        Sequence 4560, Ap
```

```
   13    506.5    20.6   1914  4  US-09-291-922-19           Sequence 19, Appl
   14      499    20.3   2592  4  US-09-591-025-8            Sequence 8, Appli
   15      498    20.3   5227  4  US-09-919-172-79           Sequence 79, Appl
   16      495    20.1   2592  4  US-09-894-927B-8           Sequence 8, Appli
   17      489    19.9   1566  4  US-09-489-039A-2378        Sequence 2378, Ap
   18      481    19.6    272  4  US-09-016-434-855          Sequence 855, App
   19      478    19.5   1689  4  US-10-162-012-45           Sequence 45, Appl
   20      478    19.5   4385  4  US-10-162-012-43           Sequence 43, Appl
   21    468.5    19.1   2089  4  US-09-291-922-25           Sequence 25, Appl
   22    468.5    19.1   3000  2  US-08-928-692-9            Sequence 9, Appli
   23    468.5    19.1   3000  4  US-09-339-972-9            Sequence 9, Appli
   24    464.5    18.9   1853  4  US-09-291-922-23           Sequence 23, Appl
   25    458.5    18.7   1872  4  US-09-291-922-27           Sequence 27, Appl
   26      453    18.4 4411529  3  US-09-103-840A-1           Sequence 1, Appli
   27      446    18.2   2017  4  US-09-291-922-21           Sequence 21, Appl
   28    439.5    17.9   1776  4  US-09-679-686B-11          Sequence 11, Appl
   29      429    17.5 4403765  3  US-09-103-840A-2           Sequence 2, Appli
   30      428    17.4    272  4  US-09-016-434-440          Sequence 440, App
   31      423    17.2    972  4  US-09-134-001C-1798        Sequence 1798, Ap
   32    398.5    16.2   2343  2  US-09-031-392-1            Sequence 1, Appli
   33    398.5    16.2   2343  3  US-09-299-549-1            Sequence 1, Appli
   34    398.5    16.2   2343  4  US-09-610-417-1            Sequence 1, Appli
c  35    382.5    15.6   3035  4  US-09-833-381-1987         Sequence 1987, Ap
   36    368.5    15.0   2842  4  US-09-620-312D-574         Sequence 574, App
   37      339    13.8   2601  4  US-09-291-922-7            Sequence 7, Appli
   38      337    13.7    918  4  US-09-107-532A-561         Sequence 561, App
   39    296.5    12.1   1487  4  US-09-291-922-13           Sequence 13, Appl
   40    293.5    11.9   2824  4  US-09-291-922-1            Sequence 1, Appli
   41    289.5    11.8   1009  4  US-09-291-922-15           Sequence 15, Appl
   42      281    11.4   1692  4  US-09-291-922-9            Sequence 9, Appli
   43      276    11.2   1896  3  US-08-501-572-6            Sequence 6, Appli
   44      276    11.2   1896  3  US-09-040-444-6            Sequence 6, Appli
   45    257.5    10.5   1341  4  US-09-543-681A-2407        Sequence 2407, Ap

                                  ALIGNMENTS

RESULT 1
US-09-489-039A-4731
; Sequence 4731, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
;  APPLICANT:  Gary Breton et. al
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
;  TITLE OF INVENTION:  PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  2709.2004001
;  CURRENT APPLICATION NUMBER: US/09/489,039A
;  CURRENT FILING DATE:  2000-01-27
;  PRIOR APPLICATION NUMBER:  US 60/117,747
;  PRIOR FILING DATE:  1999-01-29
;  NUMBER OF SEQ ID NOS:  14342
; SEQ ID NO 4731
;   LENGTH: 1545
;   TYPE: DNA
;   ORGANISM: Klebsiella pneumoniae
US-09-489-039A-4731

Alignment Scores:
Pred. No.:              5.87e-53        Length:         1545
Score:                  605.00          Matches:        156
Percent Similarity:     51.60%          Conservative:   86
Best Local Similarity:  33.26%          Mismatches:     187
Query Match:            24.62%          Indels:         40
DB:                     4               Gaps:           9

US-09-886-954A-1 (1-477) x US-09-489-039A-4731 (1-1545)

Qy         20 AlaProArgGlyArgArg---------------ValPheLeuAlaAlaPheAlaAlaAla 34
              :::||||||   :::|||               :::      :::   |||||||||
Db        154 TCGCCGCGGACGCAACGTGATACCCGGCGGATGAACTGGTTTGTTTCCATCGCCGCGGCG 213

Qy         35 LeuGlyProLeuSerPheGlyPheAlaLeuGlyTyrSerSerProAlaIleProSerLeu 54
              :::      |||   ||||||      :::|||      |||   |||:::|||   :::
```

```
Db           214 GTGGCTGGCTTGCTCTTTGGCCTGGATATCGGCGTGATATCCGGAGCGTTGCCCTTTATA 273

Qy            55 GlnArgAlaAlaProProAlaProArgLeuAspAspAlaAlaAlaSerTrpPheGlyAla 74
                                   :::   :::|||   :::            |||      :::
Db           274 ACCGACCATTTCACCTTATCCAGCCAGCTTCAGGAG------------TGGGTGGTTAGC 321

Qy            75 ValValThrLeuGlyAlaAlaAlaGlyGlyValLeuGlyGlyTrpLeuValAspArgAla 94
                    :::   ||||||||||||   |||   :::      |||||||||      |||
Db           322 AGCATGATGTTGGGGGCGGCGATAGGCGCGCTGTTTAACGGCTGGCTGTCTTTCCGCCTT 381

Qy            95 GlyArgLysLeuSerLeuLeuLeuCysSerValProPheValAlaGlyPheAlaValIle 114
                 |||||||||   ||||||:::      :::|||   ||||||||||||
Db           382 GGCCGTAAATACAGCCTGATGGCGGGGGCCGTACTCTTTGTGGCCGGCTCTATCGGATCC 441

Qy           115 ThrAlaAlaGlnAspValTrpMetLeuLeuGlyGlyArgLeuLeuThrGlyLeuAlaCys 134
                       |||      |||   :::||||||      |||::::::   |||:::|||
Db           442 GCTTTTGCCGCCAGCGTGGAGGTGCTGCTGGTGGCCCGCGTGGTGTTGGGCGTGGCAGTC 501

Qy           135 GlyValAlaSerLeuValAlaProValTyrIleSerGluIleAlaTyrProAlaValArg 154
                 |||:::||||||      ||||||:::|||:::||||||:::|||         ||||||
Db           502 GGGATTGCCTCTTATACCGCGCCGCTGTACCTCTCCGAGATGGCCAGCGAGAACGTGCGC 561

Qy           155 GlyLeuLeuGlySerCysValGlnLeuMetValValValGlyIleLeuLeuAlaTyrLeu 174
                 |||   :::   |||      ||||||||||||   :::||||||:::||||||:::|||
Db           562 GGGAAAATGATCAGTATGTATCAGCTGATGGTCACCCTTGGCATTGTGCTGGCGTTTCTT 621

Qy           175 AlaGlyTrpValLeuGluTrp------ArgTrpLeuAlaValLeuGly------CysVal 190
                 :::               :::         |||   |||:::||||||         :::
Db           622 TCCGATACCGCCTTTAGCTACAGCGGTAACTGGCGCGCCATGCTGGGCGTGCTGGCGCTG 681

Qy           191 ProProSerLeuMetLeuLeuLeuMetCysPheMetProGluThrProArgPheLeuLeu 210
                 |||      ::::::::::::|||:::   |||:::|||   :::||||||:::|||
Db           682 CCGGCGGTGATCCTGATCATTCTGGTCGTCTTTTTGCCGAACAGCCCGCGCTGGCTGGCG 741

Qy           211 ThrGlnHisArgArgGlnGluAlaMetAlaAlaLeuArgPheLeuTrpGlySerGluGln 230
                    :::   |||      ||||||         ||||||   |||      :::   :::
Db           742 GAGAAGGGACGCCATATCGAAGCGGAAGAAGTGCTGCGGATGCTGCGCGATACCTCGGAA 801

Qy           231 GlyTrpGluAspProProIleGlyAlaGluGlnSerPheHisLeu--------------- 245
                          |||                  :::|||      |||
Db           802 AAGGCGCGCGACGAGCTTAACGAGATCCGTGAGAGCCTCAAGCTGAAGCAGGGCGGTTGG 861

Qy           246 AlaLeuLeuArg---GlnProGlyIleTyrLysProPheIleIleGlyValSerLeuMet 264
                 ||||||   :::            :::   :::         :::|||:::   |||
Db           862 GCGTTGTTTAAGATCAATCGTAACGTGCGCCGGGCGGTGTTCCTGGGCATGCTGCTGCAG 921

Qy           265 AlaPheGlnGlnLeuSerGlyValAsnAlaValMetPheTyrAlaGluThrIlePheGlu 284
                 |||   ||||||   :::|||:::|||   :::|||:::||||||      ||||||:::
Db           922 GCGATGCAGCAATTCACCGGCATGAACATCATCATGTACTATGCGCCGCGTATCTTTAAA 981

Qy           285 GluAlaLysPhe---------LysAspSerSerLeuAlaSerValValValGlyValIle 301
                    |||   |||         :::         :::|||::::::|||||||||:::
Db           982 ATGGCGGGCTTTACCACTACTGAACAGCAGATGATCGCCACCCTGGTGGTGGGCCTGACC 1041

Qy           302 GlnValLeuPheThrAlaValAlaAlaLeuIleMetAspArgAlaGlyArgArgLeuLeu 321
                    :::      |||   :::|||         :::|||:::|||||||||:::
Db          1042 TTTATGTTTGCCACCTTTATTGCGGTGTTCACAGTGGATAAAGCGGGCCGCAAGCCAGCG 1101

Qy           322 LeuValLeuSerGlyValValMetValPheSerThrSerAlaPheGlyAlaTyrPheLys 341
                 |||   :::         ||||||         |||         |||      :::
Db          1102 CTGAAAATCGGCTTTAGCGTGATGGCGCTGGGCACTCTGGTGCTGGGC------TACTGC 1155

Qy           342 LeuThrGlnGlyGlyProGlyAsnSerSerHisValAlaIleSerAlaProValSerAla 361
                 |||   |||         |||
Db          1156 CTGATGCAGTTCGACAATGGCACC------------------------------------ 1179

Qy           362 GlnProValAspAlaSerValGlyLeuAlaTrpLeuAlaValGlySerMetCysLeuPhe 381
                             ||||||   ||||||:::||||||:::||||||         :::
Db          1180 ------------GCATCCAGCGGCCTCTCCTGGCTCTCCGTCGGCATGACCATGATGTGT 1227

Qy           382 IleAlaGlyPheAlaValGlyTrpGlyProIleProTrpLeuLeuMetSerGluIlePhe 401
                 |||||||||:::|||:::         |||:::   |||:::|||   |||||||||
Db          1228 ATTGCCGGGTATGCGATGAGCGCGGCGCCGGTGGTGTGGATCCTCTGCTCCGAGATCCAG 1287

Qy           402 ProLeuHisValLysGlyValAlaThrGlyIleCysValLeuThrAsnTrpLeuMetAla 421
                 ||||||      :::                           |||||||||:::
Db          1288 CCGCTAAAATGCCGCGACTTCGGTATCACCTGCTCGACCACCACCAACTGGGTGTCGAAC 1347

Qy           422 PheLeuValThrLysGluPheSerSerLeuMetGluValLeuArgProTyrGlyAlaPhe 441
                    ::::::         |||   :::|||::::::   :::         |||   |||
Db          1348 ATGATCATCGGCGCCACTTTCCTGACGCTGCTTGACGCGATTGGCGCCGCCGGCACCTTC 1407

Qy           442 TrpLeuAlaSerAlaPheCysIlePheSerValLeuPheThrLeuPheCysValProGlu 461
                 ||||||   :::|||      :::      :::      |||   :::   :::||||||
Db          1408 TGGCTCTACACGGCGCTCAACGTGGCCTTTATCGGCATCACCTTCTGGCTGATCCCGGAA 1467

Qy           462 ThrLysGlyLysThrLeuGluGlnIle 470
                 ||||||      |||||||||   |||
Db          1468 ACCAAGAATGTCACCCTCGAGCACATT 1494

RESULT 2
US-09-489-039A-4762
; Sequence 4762, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
;  APPLICANT:  Gary Breton et. al
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
;  TITLE OF INVENTION:  PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  2709.2004001
;  CURRENT APPLICATION NUMBER: US/09/489,039A
;  CURRENT FILING DATE:  2000-01-27
;  PRIOR APPLICATION NUMBER:  US 60/117,747
;  PRIOR FILING DATE:  1999-01-29
;  NUMBER OF SEQ ID NOS:  14342
; SEQ ID NO 4762
;   LENGTH: 1431
;   TYPE: DNA
;   ORGANISM: Klebsiella pneumoniae
US-09-489-039A-4762

Alignment Scores:
Pred. No.:              2.92e-50       Length:        1431
Score:                  578.50         Matches:       149
Percent Similarity:     48.74%         Conservative:  84
Best Local Similarity:  31.17%         Mismatches:    180
Query Match:            23.54%         Indels:        65
DB:                     4              Gaps:          10

US-09-886-954A-1 (1-477) x US-09-489-039A-4762 (1-1431)

Qy            27 PheLeuAlaAlaPheAlaAlaAlaLeuGlyProLeuSerPheGlyPheAlaLeuGlyTyr 46
                 |||         |||   |||||||||      |||   ||||||      :::|||
Db            82 TTCTTCGTCTGTTTCCTCGCCGCGCTGGCTGGCCTGCTGTTCGGCCTTGATATCGGTGTT 141

Qy            47 SerSerProAlaIleProSerLeuGlnArgAlaAlaProProAlaProArgLeuAspAsp 66
                    :::   |||:::|||   :::                        ::::::
Db           142 ATTGCGGGTGCCTTACCCTTTATTGCCAATGAGTTC------------CAGATTTCCGCC 189

Qy            67 AlaAlaAlaSerTrpPheGlyAlaValValThrLeuGlyAlaAlaAlaGlyGlyValLeu 86
                             |||      :::   :::      |||||||||   |||   |||
Db           190 CACACCCAGGAGTGGGTGGTCAGCTCCATGATGTTCGGGGCTGCCGTCGGCGCGGTCGGC 249

Qy            87 GlyGlyTrpLeuValAspArgAlaGlyArgLysLeuSerLeuLeuLeuCysSerValPro 106
                    |||||||||      :::   |||||||||   ||||||::::::   ::::::
Db           250 AGCGGCTGGCTCTCTTTCAAACTGGGCCGGAAAAAGAGCCTGATGATCGGCGCCATCCTC 309

Qy           107 PheValAlaGlyPheAlaValIleThrAlaAlaGlnAspValTrpMetLeuLeuGlyGly 126
                 ||||||||||||               ||||||   :::|||   :::||||||
Db           310 TTCGTCGCCGGTTCGCTGTTCTCTGCCGCCGCGCCAAACGTCGAGATCCTGCTGGTTTCC 369

Qy           127 ArgLeuLeuThrGlyLeuAlaCysGlyValAlaSerLeuValAlaProValTyrIleSer 146
                 |||:::|||   |||||||||   ||||||||||||      ||||||:::|||:::|||
Db           370 CGTGTGCTGCTCGGCCTGGCGGTGGGCGTCGCCTCATATACGGCTCCGCTGTATCTGTCG 429

Qy           147 GluIleAlaTyrProAlaValArgGlyLeuLeuGlySerCysValGlnLeuMetValVal 166
                 |||||||||         :::||||||   :::   |||      |||||||||:::
```

```
Db         430 GAAATCGCCCCAGAAAAAATTCGCGGCAGTATGATTTCCATGTACCAGCTGATGATCACC 489

Qy         167 ValGlyIleLeuLeuAlaTyrLeuAlaGlyTrpValLeuGlu------------TrpArg 182
               :::|||||||||   |||||||||:::                           ||||||
Db         490 ATCGGGATCCTTGGCGCCTATCTCTCTGACACCGCTTTCAGCTACAGCGGCGCATGGCGC 549

Qy         183 TrpLeuAlaValLeuGlyCysValProProSerLeuMetLeuLeuLeuMetCysPheMet 202
               |||:::      :::      :::|||      |||:::|||:::   :::   |||:::
Db         550 TGGATGCTCGGGGTTATCATCATTCCGGCGGTTTTGCTGCTGATCGGCGTTATCTTCCTG 609

Qy         203 ProGluThrProArgPheLeuLeuThrGlnHisArgArgGlnGluAlaMetAlaAlaLeu 222
               |||::::::||||||:::         :::   |||      :::|||         |||
Db         610 CCGGACAGCCCGCGCTGGTTCGCCGCCAAACGTCGCTTTGTCGATGCGGAACGCGTGCTG 669

Qy         223 ------------------------------------------------------ArgPhe 224

Db         670 CTGCGCCTGCGCGATACCAGCGCCGAAGCGAAACGCGAGCTCGATGAAATCCGTGAAAGC 729

Qy         225 LeuTrpGlySerGluGlnGlyTrpGluAspProProIleGlyAlaGluGlnSerPheHis 244
               |||         :::   ||||||
Db         730 CTGAAGGTAAAACAGTCCGGCTGG------------------------------------ 753

Qy         245 LeuAlaLeuLeuArgGlnProGlyIleTyrLysProPheIle---IleGlyValSerLeu 263
                  :::|||   :::            ::::::      :::   :::|||:::   |||
Db         754 ---TCGCTGTTTAAAGACAACAGCAACTTCCGCCGCGCGGTGTTCCTCGGCATCCTGCTG 810

Qy         264 MetAlaPheGlnGlnLeuSerGlyValAsnAlaValMetPheTyrAlaGluThrIlePhe 283
                        ||||||   :::|||:::|||   :::|||:::||||||      ||||||
Db         811 CAGGTGATGCAACAGTTCACCGGGATGAACGTCATCATGTACTACGCGCCGAAGATCTTT 870

Qy         284 GluGluAlaLysPheLysAspSerSer---------LeuAlaSerValValValGlyVal 300
               |||   |||   :::   :::::::::               :::|||:::||||||:::
Db         871 GAGCTGGCGGGTTATGCCAACACCACTGAGCAAATGTGGGGGACAGTGATCGTCGGTCTG 930

Qy         301 IleGlnValLeuPheThrAlaValAlaAlaLeuIleMetAspArgAlaGlyArgArgLeu 320
                     ||||||   |||   :::|||      ::::::||||||   ||||||:::
Db         931 ACTAACGTGCTGGCCACCTTTATCGCCATCGGTCTGGTCGACCGCTGGGGCCGTAAACCG 990

Qy         321 LeuLeuValLeuSerGlyValValMetValPheSerThrSerAlaPheGlyAlaTyrPhe 340
                  |||:::|||      :::||||||               |||   |||
Db         991 ACGCTGATCCTTGGCTTTATCGTGATG---------------GCCGCGGGA--------- 1026

Qy         341 LysLeuThrGlnGlyGlyProGlyAsnSerSerHisValAlaIleSerAlaProValSer 360
                           |||      |||         |||:::   |||   :::      :::
Db        1027 ---------ATGGGCGTCCTGGGTACCATGATGCACATCGGCATTCACTCCTCTACCGCC 1077

Qy         361 AlaGlnProValAspAlaSerValGlyLeuAlaTrpLeuAlaValGlySerMetCysLeu 380
                                                ::::::||||||      :::   :::
Db        1078 ------------------------------CAGTACATCGCCGTCCTGATGCTGCTGATG 1107

Qy         381 PheIleAlaGlyPheAlaValGlyTrpGlyProIleProTrpLeuLeuMetSerGluIle 400
               ||||||   |||||||||:::      ||||||:::   |||:::|||   |||||||||
Db        1108 TTCATCGTCGGTTTCGCCATGAGCGCCGGCCCACTGATTTGGGTACTGTGCTCCGAAATC 1167

Qy         401 PheProLeuHisValLysGlyValAlaThrGlyIleCysValLeuThrAsnTrpLeuMet 420
                  ||||||      :::                           |||||||||:::
Db        1168 CAGCCGCTGAAAGGCCGCGACTTCGGTATCACCTGCTCCACAGCGACCAACTGGATTGCC 1227

Qy         421 AlaPheLeuValThrLysGluPheSerSerLeuMetGluValLeuArgProTyrGlyAla 440
                     :::|||         |||   :::::::::      |||
Db        1228 AACATGATTGTCGGCGCCACCTTCCTGACCATGCTCAACTCGCTGGGCAGCGCCAATACC 1287

Qy         441 PheTrpLeuAlaSerAlaPheCysIlePheSerValLeuPheThrLeuPheCysValPro 460
               ||||||:::               :::      :::|||   ||||||:::   :::|||
Db        1288 TTCTGGGTGTACGGCGGTCTGAACGTGCTGTTTATCCTGCTGACGCTGTGGCTGATCCCG 1347

Qy         461 GluThrLysGlyLysThrLeuGluGlnIleThrAlaHis---PheGluGlyArg 477
               |||||||||      :::||||||   |||      :::      :::||||||
Db        1348 GAAACCAAAAACGTCTCGCTGGAACATATTGAACGTAACCTGATGCAGGGTCGT 1401

RESULT 3
US-09-023-655-1104
; Sequence 1104, Application US/09023655
; Patent No. 6607879
;  GENERAL INFORMATION:
;    APPLICANT:  Cocks, Benjamin G.
;    APPLICANT:  Susan G. Stuart
;    APPLICANT:  Jeffrey J. Seilhamer
;    TITLE OF INVENTION:  COMPOSITION FOR THE DETECTION OF BLOOD CELL GENE
;    TITLE OF INVENTION:  EXPRESSION
;    NUMBER OF SEQUENCES:  1508
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  INCYTE PHARMACEUTICALS, INC.
;      STREET:  3174 PORTER DRIVE
;      CITY:  PALO ALTO
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Word Perfect 6.1 for Windows/MS-DOS 6.2
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/023,655
;      FILING DATE:  HEREWITH
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Zeller, Karen J.
;      REGISTRATION NUMBER:  37,071
;      REFERENCE/DOCKET NUMBER:  PA-0001 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (650) 855-0555
;      TELEFAX:  (650) 845-4166
;  INFORMATION FOR SEQ ID NO:  1104:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3915 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  GENBANK
;      CLONE:  g183684
US-09-023-655-1104

Alignment Scores:
Pred. No.:               9.58e-49       Length:          3915
Score:                   570.50         Matches:         152
Percent Similarity:      49.90%         Conservative:    88
Best Local Similarity:   31.60%         Mismatches:      190
Query Match:             23.22%         Indels:          51
DB:                      4              Gaps:            9

US-09-886-954A-1 (1-477) x US-09-023-655-1104 (1-3915)

Qy          26 ValPheLeuAlaAlaPheAlaAlaAlaLeuGlyProLeuSerPheGlyPheAlaLeuGly 45
               :::      |||         |||   :::|||         ||||||:::      |||
Db         270 CTGATATTTGCCATCACAGTTGCTACAATCGGCTCTTTCCAATTTGGCTACAACACTGGG 329

Qy          46 TyrSerSerProAlaIleProSerLeuGlnArg--------------------------- 56
                     :::               ::::::
Db         330 GTCATCAATGCTCCTGAGAAGATCATAAAGGAATTTATCAATAAAACTTTGACGGACAAG 389

Qy          57 ---AlaAlaProProAlaProArgLeuAspAspAlaAlaAlaSerTrpPheGlyAlaVal 75
                     |||||||||:::      |||      :::      |||         |||:::
Db         390 GGAAATGCCCCACCCTCTGAGGTGCTGCTCACGTCTCTCTGGTCCTTGTCTGTGGCCATA 449

Qy          76 ValThrLeuGlyAlaAlaAlaGlyGlyValLeuGlyGlyTrpLeuValAspArgAlaGly 95
                  ::::::|||         |||            |||      |||:::|||   |||
Db         450 TTTTCCGTCGGGGGTATGATCGGCTCCTTTTCCGTCGGACTCTTCGTCAACCGCTTTGGC 509

Qy          96 ArgLysLeuSerLeuLeuLeuCysSerValProPheValAlaGly---------PheAla 112
```

```
              |||:::   |||:::|||:::   ::::::      |||   |||
Db        510 AGGCGCAATTCAATGCTGATTGTCAACCTGTTGGCTGTCACTGGTGGCTGCTTTATGGGA 569

Qy        113 ValIleThrAlaAlaGlnAspValTrpMetLeuLeuGlyGlyArgLeuLeuThrGlyLeu 132
              :::         |||:::   |||   ||||||:::   |||||||||:::   ||||||
Db        570 CTGTGTAAAGTAGCTAAGTCGGTTGAAATGCTGATCCTGGGTCGCTTGGTTATTGGCCTC 629

Qy        133 AlaCysGlyValAlaSerLeuValAlaProValTyrIleSerGluIleAlaTyrProAla 152
                 ||||||:::   :::         |||:::||||||   ||||||:::      |||
Db        630 TTCTGCGGACTCTGCACAGGTTTTGTGCCCATGTACATTGGAGAGATCTCGCCTACTGCC 689

Qy        153 ValArgGlyLeuLeuGlySerCysValGlnLeuMetValValValGlyIleLeuLeuAla 172
              :::||||||      |||:::      ||||||   :::|||||||||||||||:::|||
Db        690 CTGCGGGGTGCCTTTGGCACTCTCAACCAGCTGGGCATCGTTGTTGGAATTCTGGTGGCC 749

Qy        173 TyrLeuAlaGly------TrpValLeu------GluTrpArgTrpLeuAlaValLeuGly 188
                 :::   |||      ::::::|||      |||      |||      :::||||||
Db        750 CAGATCTTTGGTCTGGAATTCATCCTTGGGTCTGAAGAGCTATGGCCGCTGCTACTGGGT 809

Qy        189 ------CysValProProSerLeuMetLeuLeuLeuMetCysPheMetProGluThrPro 206
                       :::|||      |||            :::   |||   ||||||:::|||
Db        810 TTTACCATCCTTCCTGCTATCCTACAAAGTGCAGCCCTTCCATTTGCCCTGAAAGTCCC 869

Qy        207 ArgPheLeuLeuThrGlnHisArgArgGlnGlu---AlaMetAlaAlaLeuArgPheLeu 225
              ||||||||||||         :::   :::|||   |||         |||:::   |||
Db        870 AGATTTTTGCTCATTAACAGAAAAGAAGAGGAGAATGCTAAGCAGATCCTCCAGCGGTTG 929

Qy        226 TrpGlySerGluGlnGlyTrpGluAspProProIleGlyAlaGluGlnSerPheHis--- 244
              ||||||::::::         :::|||               ::::::|||
Db        930 TGGGGCACCCAGGATGTATCCCAAGACATCCAGGAGATGAAAGATGAGAGTGCAAGGATG 989

Qy        245 ------------------------LeuAlaLeuLeuArgGlnProGlyIleTyrLysPro 256
                                      |||   |||   |||               :::|||
Db        990 TCACAAGAAAAGCAAGTCACCGTGCTAGAGCTCTTTAGAGTGTCCAGCTACCGACAGCCC 1049

Qy        257 PheIleIleGlyValSerLeuMetAlaPheGlnGlnLeuSerGlyValAsnAlaValMet 276
                 ||||||   :::   |||         |||||||||||||||:::|||||||||
Db       1050 ATCATCATTTCCATTGTGCTCCAGCTCTCTCAGCAGCTCTCTGGGATCAATGCTGTGTTC 1109

Qy        277 PheTyrAlaGluThrIlePheGluGluAlaLysPheLysAspSerSerLeuAlaSerVal 296
              :::|||:::      ||||||::::::|||      ::::::         |||::::::
Db       1110 TATTACTCAACAGGAATCTTCAAGGATGCAGGTGTTCAAGAGCCCATCTATGCCACCATC 1169

Qy        297 ValValGlyValIleGlnValLeuPheThrAlaValAlaAlaLeuIleMetAspArgAla 316
                    ||||||:::      :::||||||   |||:::      :::::::::||||||
Db       1170 GGCGCGGGTGTGGTTAATACTATCTTCACTGTAGTTTCTCTATTTCTGGTGGAAAGGGCA 1229

Qy        317 GlyArgArgLeuLeuLeuValLeuSerGlyValValMetValPheSerThrSerAlaPhe 336
              |||||||||   |||   ::::::
Db       1230 GGAAGAAGGACTCTGCATATGATA------------------------------------ 1253

Qy        337 GlyAlaTyrPheLysLeuThrGlnGlyGlyProGlyAsnSerSerHisValAlaIleSer 356
                                         |||   |||      :::      :::
Db       1254 ---------------------------GGCCTTGGAGGGATGGCTTTTTGTTCCACGCTC 1286

Qy        357 AlaProValSerAlaGlnProValAspAlaSerValGlyLeuAlaTrpLeuAlaValGly 376
                    ||||||            |||         |||::::::::::::   :::|||
Db       1287 ATGACTGTTTCTTTGTTATTAAAGGATAACTATAATGGGATGAGCTTTGTCTGTATTGGG 1346

Qy        377 SerMetCysLeuPheIleAlaGlyPheAlaValGlyTrpGlyProIleProTrpLeuLeu 396
              ::::::   :::|||:::|||   |||   :::|||   |||||||||||||||   :::
Db       1347 GCTATCTTGGTCTTTGTAGCCTTCTTTGAAATTGGACCAGGCCCCATTCCCTGGTTTATT 1406

Qy        397 MetSerGluIlePheProLeuHisValLysGlyValAlaThrGlyIleCysValLeuThr 416
              ::::::|||:::|||            :::      |||      :::         :::
Db       1407 GTGGCCGAACTCTTCAGCCAGGGCCCCCGCCCAGCTGCGATGGCAGTGGCCGGCTGCTCC 1466

Qy        417 AsnTrpLeuMetAlaPheLeuValThrLysGluPheSerSerLeuMetGluValLeuArg 436
              ||||||         |||||||||         |||   |||            |||
Db       1467 AACTGGACCTCCAACTTCCTAGTCGGATTGCTCTTCCCCTCCGCTGCTCACTATTTAGGA 1526

Qy        437 ProTyrGlyAlaPheTrpLeuAlaSerAlaPheCysIlePheSerValLeuPheThrLeu 456
                 |||      |||   :::   :::   |||   |||      :::   ||||||
Db       1527 GCCTAC---GTTTTTATTATCTTCACCGGCTTCCTCATTACCTTCTTGGCTTTTACCTTC 1583

Qy        457 PheCysValProGluThrLysGlyLysThrLeuGluGlnIleThrAlaHisPheGluGly 476
              |||   ||||||||||||:::|||:::|||   |||   ||||||      |||||||||
Db       1584 TTCAAAGTCCCTGAGACCCGTGGCAGGACTTTTGAGGATATCACACGGGCCTTTGAAGGG 1643

Qy        477 Arg 477
              :::
Db       1644 CAG 1646

RESULT 4
US-09-679-686B-17
; Sequence 17, Application US/09679686B
; Patent No. 6624343
; GENERAL INFORMATION:
;  APPLICANT: Allen, Stephen M.
;  APPLICANT:  Lightner, Jonathan E.
;  APPLICANT:  Rafalski, J. Antoni
;  APPLICANT:  Thorpe, Catherine J.
;  TITLE OF INVENTION: HEXOSE CARRIER PROTEINS
;  FILE REFERENCE: BB1160 US NA
;  CURRENT APPLICATION NUMBER: US/09/679,686B
;  CURRENT FILING DATE:  2003-01-16
;  PRIOR APPLICATION NUMBER: 60/081,131
;  PRIOR FILING DATE: 1998-04-09
;  PRIOR APPLICATION NUMBER: PCT/US99/07561
;  PRIOR FILING DATE: 1999-04-07
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 17
;   LENGTH: 1752
;   TYPE: DNA
;   ORGANISM: Triticum aestivum
US-09-679-686B-17

Alignment Scores:
Pred. No.:               2.17e-46        Length:          1752
Score:                   542.50          Matches:         150
Percent Similarity:      47.50%          Conservative:    88
Best Local Similarity:   29.94%          Mismatches:      202
Query Match:             22.08%          Indels:          61
DB:                      4               Gaps:            14

US-09-886-954A-1 (1-477) x US-09-679-686B-17 (1-1752)

Qy         17 GlyGlySerAlaProArgGlyArgArg---------ValPheLeuAlaAlaPheAlaAla 33
              ||||||            ||||||            ||||||               |||
Db         96 GGCGGCAAGGACTACCCTGGCAGGCTCACCCTCTTCGTCTTCTTCACCTGCGTCGTCGCC 155

Qy         34 AlaLeuGlyProLeuSerPheGlyPheAlaLeuGlyTyrSerSerProAlaIle------ 51
              |||   |||   |||   ||||||:::   :::|||   |||
Db        156 GCCACCGGCGGCCTCATCTTTGGATATGACATCGGTATCTCAGGGGGCGTTACGTCCATG 215

Qy         52 ---ProSerLeuGlnArgAlaAlaProProAla--------------------------- 61
                 |||   |||::::::      |||
Db        216 AACCCTTTCCTGAAAAAGTTCTTCCCGGAGGTGTATGACAAGAAGCAGATGAAGGGCTCC 275

Qy         62 ------------ProArgLeuAspAspAlaAlaAlaSerTrpPheGlyAlaValValThr 77
                             :::   |||:::               |||   :::   :::
Db        276 GCCAGCCAGTACTGCAAGTACGACAACCAGCTGCTCCAGACCTTCACCTCCTCCCTCTAC 335

Qy         78 LeuGlyAlaAlaAlaGlyGlyValLeuGlyGlyTrpLeuValAspArgAlaGlyArgLys 97
              |||   |||                           :::            |||||||||
Db        336 CTCGCGGCGCTCGTCTCCTCCTTCTTCGCCGCCACCGTCACACGTGTCGTGGGCCGTAAG 395

Qy         98 LeuSerLeuLeuLeuCysSerValProPheValAlaGlyPheAlaValIleThrAlaAla 117
                 |||:::            :::   |||:::   |||   |||:::      ||||||
Db        396 TGGTCCATGTTCACCGGAGGGCTCACCTTCCTCATCGGCGCTGCGCTTAACGGGGCGGCG 455

Qy        118 GlnAspValTrpMetLeuLeuGlyGlyArgLeuLeuThrGlyLeuAlaCysGlyValAla 137
              :::::::::   ||||||:::   ||||||:::|||   |||:::      |||   |||
Db        456 GAGAACATCGCCATGCTCATCGTCGGACGCATCCTCCTCGGTGTCGGCGTTGGCTTCGCC 515
```

```
Qy      138 SerLeuValAlaProValTyrIleSerGluIleAlaTyrProAlaValArgGlyLeuLeu 157
            :::         |||||||||:::||||||:::|||         :::||||||:::|||
Db      516 AATCAGTCTGTGCCGGTGTACCTGTCGGAGATGGCGCCTGCGCGTCTCCGGGGCATGCTC 575

Qy      158 GlySerCysValGlnLeuMetValValValGlyIleLeuLeuAlaTyrLeu--------- 174
                        |||||||||:::   :::|||||||||   |||   |||
Db      576 AACATCGGGTTCCAGCTCATGATCACCATCGGCATCCTGGCGGCGCTCATCAATTAC 635

Qy      175 ------------------AlaGlyTrpValLeuGluTrpArgTrpLeuAlaValLeuGly 188
                              ||||||:::         ||||||            :::
Db      636 GACACCAACAAGATCAAGGCCGGGTAC------GGGTGGCGCATCAGCCTGGCCATCGCG 689

Qy      189 CysValProProSerLeuMetLeuLeuLeuMetCysPheMetProGluThrProArgPhe 208
               ||||||      ::::::   |||         |||:::|||:::||||||
Db      690 GCCGTCCCGGCGGGCATCATCACCCTGGGGTCCTTTTTCCTCCCCGACACCCCCAACTCC 749

Qy      209 LeuLeuThrGlnHisArgArgGlnGluAlaMetAlaAlaLeuArgPheLeuTrpGlySer 228
            |||:::   :::         :::   |||         |||      :::   ||||||
Db      750 CTCATCGAGCGTGGCCACCCGGAGGCGGCGCGCCGCATGCTCAACCGCATCCGCGGCAGC 809

Qy      229 Glu------------GlnGlyTrpGluAspProProIleGlyAlaGluGlnSerPheHis 244
            :::            :::   :::   |||      :::   :::|||:::|||
Db      810 GACGTGGACATCAGCGAGGAGTACGCGGACCTGGTGGTGGCGAGCGAGGAGTCC------ 863

Qy      245 LeuAlaLeuLeuArgGlnPro------------------GlyIleTyrLysPro---Phe 257
                  |||::::::   |||                        |||:::|||
Db      864 ---AAGCTGGTGCAGCACCCGTGGCGCAACATCCTGCAGCGCAAGTACCGGCCCCAGCTG 920

Qy      258 IleIleGlyValSerLeuMetAlaPheGlnGlnLeuSerGlyValAsnAlaValMetPhe 277
               :::   :::   :::      ||||||||||||:::|||:::|||   :::||||||
Db      921 ACCATGGCGATCATGATCCCCTTCTTCCAGCAGCTGACGGGCATCAACGTCATCATGTTC 980

Qy      278 TyrAlaGluThrIlePheGluGluAlaLysPheLys---AspSerSerLeuAlaSer--- 295
            ||||||      :::||||||         ||||||   |||:::||||||   |||
Db      981 TACGCGCCGGTGCTGTTCGAGACGCTGGGGTTCAAGGGCGACGCGTCGCTCATGTCGGCC 1040

Qy      296 ValValValGlyValIleGlnValLeuPheThrAlaValAlaAlaLeuIleMetAspArg 315
            |||:::   |||::::::   |||      |||   |||:::         :::||||||
Db     1041 GTCATCACGGGCCTGGTCAACGTGTTCGCGACGCTCGTGTCCGTGTTCACCGTCGACCGG 1100

Qy      316 AlaGlyArgArgLeuLeuLeuValLeuSerGlyValValMetValPheSerThrSerAla 335
               |||||||||   |||   :::      |||      |||:::   |||
Db     1101 CTGGGTCGCCGGAAGCTGTTCCTGCAGGGCGGCACGCAGATGCTGCTGAGCCAGCTGGTG 1160

Qy      336 PheGlyAlaTyrPheLysLeuThrGlnGlyGlyProGlyAsnSerSerHisValAlaIle 355
               |||            :::      |||      |||
Db     1161 GTGGGCACCCTGATCGCGGTCAAGTTCGGGACGAGCGGC--------------------- 1199

Qy      356 SerAlaProValSerAlaGlnProValAspAlaSerValGlyLeuAlaTrpLeuAlaVal 375
                     |||         |||               |||   |||         |||
Db     1200 ---------GTGGGGGAGATGCCC------------AAGGGGTACGCGGCGGCGGTGGTG 1238

Qy      376 GlySerMetCysLeuPheIleAlaGlyPheAlaValGlyTrpGlyProIleProTrpLeu 395
                  :::||||||::::::||||||||||||      |||||||||:::   ||||||
Db     1239 CTCTTCATCTGCCTCTATGTGGCCGGGTTCGCGTGGTCGTGGGGGCCCCTGGGGTGGCTG 1298

Qy      396 LeuMetSerGluIlePheProLeuHisValLysGlyValAlaThrGlyIleCysValLeu 415
            :::   ||||||||||||||||||   ::::::               |||   |||
Db     1299 GTGCCCAGCGAGATCTTCCCGCTGGAGATCAGGCCGGCGGGGCAGAGCATCAACGTGTCG 1358

Qy      416 ThrAsnTrpLeuMetAlaPheLeuValThrLysGluPheSerSerLeuMetGluValLeu 435
               |||   |||      |||::::::   :::   |||   :::::::::      :::
Db     1359 GTGAACATGCTCTTCACCTTCGTCATCGCGCAGGCGTTCCTCACCATGCTCTGCCACATG 1418

Qy      436 ArgProTyrGlyAlaPheTrpLeuAlaSerAlaPheCysIlePheSerValLeuPheThr 455
            :::   :::|||   |||:::      :::   :::   :::         :::|||
Db     1419 AAG---TTCGGCCTCTTCTACTTCTTCGCCGGCTGGGTGGTGATCATGACCGTCTTCATC 1475

Qy      456 LeuPheCysValProGluThrLysGlyLysThrLeuGluGlnIleThrAlaHisPheGlu 475
                     :::||||||||||||         :::|||::::::         ::::::
Db     1476 GCGCTCTTCCTGCCGGAGACCAAGAACGTGCCCATCGAGGAGATGGTGCTCGTCTGGAAG 1535

Qy      476 Gly 476

            |||
Db     1536 GGA 1538

RESULT 5
US-09-679-686B-1
; Sequence 1, Application US/09679686B
; Patent No. 6624343
; GENERAL INFORMATION:
;  APPLICANT: Allen, Stephen M.
;  APPLICANT:  Lightner, Jonathan E.
;  APPLICANT:  Rafalski, J. Antoni
;  APPLICANT:  Thorpe, Catherine J.
;  TITLE OF INVENTION: HEXOSE CARRIER PROTEINS
;  FILE REFERENCE: BB1160 US NA
;  CURRENT APPLICATION NUMBER: US/09/679,686B
;  CURRENT FILING DATE:  2003-01-16
;  PRIOR APPLICATION NUMBER: 60/081,131
;  PRIOR FILING DATE: 1998-04-09
;  PRIOR APPLICATION NUMBER: PCT/US99/07561
;  PRIOR FILING DATE: 1999-04-07
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 1
;   LENGTH: 1675
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1624)
;   OTHER INFORMATION: n = a, c, g or t
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1638)
;   OTHER INFORMATION: n = a, c, g or t
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1668)
;   OTHER INFORMATION: n = a, c, g or t
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1670)
;   OTHER INFORMATION: n = a, c, g or t
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1674)
;   OTHER INFORMATION: n = a, c, g or t
US-09-679-686B-1

Alignment Scores:
Pred. No.:                2.21e-45      Length:         1675
Score:                    532.50        Matches:        147
Percent Similarity:       44.89%        Conservative:   90
Best Local Similarity:    27.84%        Mismatches:     190
Query Match:              21.67%        Indels:         101
DB:                       4             Gaps:           13

US-09-886-954A-1 (1-477) x US-09-679-686B-1 (1-1675)

Qy        7 GluGluThrGlnProLeuLeuGlyProProGlyGlySerAlaProArgGlyArgArg--- 25
            |||:::::::::         |||      ||||||   |||      ||||||
Db      150 GAGGACAGCGAGATGGCCGGCGGGTTTGGCGGCGGGGCAGCCGCCGGGAGGGCCGAG 209

Qy       26 ---------------------------ValPheLeuAlaAlaPheAlaAlaAlaLeuGly 36
                                             ||||||            :::   |||
Db      210 CTCTACGAGGGCAAGATCACCGGCTACTTCATCCTCGCTTGCATCGTCGGCTCCTTCGGC 269

Qy       37 ProLeuSerPheGlyPheAlaLeuGlyTyrSerSerProAlaIleProSerLeuGlnArg 56
                     ||||||:::   ||||||   ||||||
Db      270 GGATCCCTCTTCGGCTATGACCTCGGAGTCTCCAGC------------------------ 305

Qy       57 AlaAlaProProAlaProArgLeuAspAspAlaAlaAlaSerTrpPhe------------ 72
                                 :::||||||            :::|||
Db      306 ---------GGCGTGACTTCCATGGACGACTTCCTGGTGAAGTTCTTCCCGGACGTGTAC 356
```

```
Qy      72 ------------------------------------------------------------ 72

Db     357 CGGCGGAAGCAGGCGCACCTGCACGAGACGGACTACTGCAAGTACGACAACCAGGTGCTG 416

Qy      73 ------------------------------GlyAlaValValThrLeuGlyAlaAlaAla 82
                                          |||   |||   |||   ||||||:::
Db     417 ACGCTCTTCACCTCGTCGCTCTACTTCGCGGGCCTCGTCTCCACCTTCGGCGCCTCC--- 473

Qy      83 GlyGlyValLeuGlyGlyTrpLeuValAspArgAlaGlyArgLysLeuSerLeuLeuLeu 102
                             ::::::      |||   ||||||:::   |||::::::
Db     474 ------------------TACGTGACCAAGCGCCACGGCCGGCGCGCCAGCATCATGGGT 515

Qy     103 CysSerValProPheValAlaGlyPheAlaValIleThrAlaAlaGlnAspValTrpMet 122
              :::      |||      |||   |||:::      ||||||   ::::::   |||
Db     516 GGCGCCGCCAGCTTCTTCCTCGGCGGCGCCATCAACGGCGCCGCCATGAACATCGCCATG 575

Qy     123 LeuLeuGlyGlyArgLeuLeuThrGlyLeuAlaCysGlyValAlaSerLeuValAlaPro 142
           |||:::   ||||||:::|||   |||:::      |||   |||:::         |||
Db     576 CTCATCGTCGGACGCATCCTCCTCGGCGTCGGCGTCGGCTTCGCAAATCAGGCCGTGCCT 635

Qy     143 ValTyrIleSerGluIleAlaTyrProAlaValArgGlyLeuLeuGlySerCysValGln 162
           ||||||:::||||||:::|||         :::||||||:::|||            |||
Db     636 GTGTACCTGTCGGAGATGGCGCCGGCGCGTCTCCGGGGCATGCTCAACATCGGCTTCCAG 695

Qy     163 LeuMetValValValGlyIleLeuLeuAlaTyrLeu------------------------ 174
           ||||||:::   :::|||||||||   |||   |||
Db     696 CTGATGATCACCATCGGCATCCTGGCGGCGGAGCTCATCAACTACGGCACCAACAAGATC 755

Qy     175 ---AlaGlyTrpValLeuGluTrpArgTrpLeuAlaValLeuGlyCysValProProSer 193
              ||||||:::         ||||||            |||      ||||||   :::
Db     756 AAGGCCGGGTAC------GGGTGGCGCGTGAGCCTGGCGCTGGCGGCGGTGCCGGCGGCC 809

Qy     194 LeuMetLeuLeuLeuMetCysPheMetProGluThrProArgPheLeuLeuThrGlnHis 213
           ::::::   |||         |||:::|||:::||||||      ||||||   :::
Db     810 ATCATCACCCTGGGCTCCCTCTTCCTCCCGGACACCCCCAACTCGCTGCTGGAGCGGGGC 869

Qy     214 ArgArgGlnGluAlaMetAlaAlaLeuArgPheLeuTrpGlySerGlu---------Gln 230
                 :::||||||         ||||||   :::   |||::::::         :::
Db     870 CACCCGGAGGAGGCACGCCGCATGCTCCGCCGCATCCGCGGCACGGACGACATCGGCGAG 929

Qy     231 GlyTrpGluAspProProIleGlyAlaGluGlnSerPheHisLeuAla------------ 246
              :::   |||            :::|||::::::      :::
Db     930 GAGTACGCGGACCTGGTGGCGGCCAGCGAGGAGGCCCGCCAGGTGCGCCACCCGTGGCGG 989

Qy     247 ---LeuLeuArgGlnProGlyIleTyrLys---ProPheIleIleGlyValSerLeuMet 264
              :::||||||:::         |||:::            :::   |||::::::
Db     990 AACATCCTGCGCCGCCGG------TACCGCGCGCAGCTCACCATGGCCGTCGCGATCCCC 1043

Qy     265 AlaPheGlnGlnLeuSerGlyValAsnAlaValMetPheTyrAlaGluThrIlePheGlu 284
              ||||||||||||   |||:::|||   :::||||||||||||      :::|||:::
Db    1044 TTCTTCCAGCAGCTCAAGGGGATCAACGTCATCATGTTCTACGCGCCCGTGCTGTTCGAC 1103

Qy     285 GluAlaLysPheLysAspSerSer------LeuAlaSerValValValGlyValIleGln 302
                    ||||||      :::      ::::::||||||:::   |||::::::
Db    1104 ACGCTGGGATTCAAGAAAGAAGCCTTCCTCATGTCCTCCGTCATCACGGGCCTCGTCAAC 1163

Qy     303 ValLeuPheThrAlaValAlaAlaLeuIleMetAspArgAlaGlyArgArgLeuLeuLeu 322
           |||      |||   |||:::   :::   :::||||||   |||||||||   |||
Db    1164 GTCTTCGCCACCGTCGTGTCCATCGTCACCGTCGACCGCGTCGGCCGCCGCAAGCTGTTC 1223

Qy     323 ValLeuSerGlyValValMetValPheSerThrSerAlaPheGlyAlaTyrPheLysLeu 342
           :::      |||      |||:::                  |||
Db    1224 CTCCAGGGCGGCGCGCAGATGATCGTGTGCCAGCTCATCGTGGGCACGCTCATCGCCGCC 1283

Qy     343 ThrGlnGlyGlyProGlyAsnSerSerHisValAlaIleSerAlaProValSerAlaGln 362
                 |||      |||
Db    1284 AAGTTCGGGACCAGCGGCACGGGG------------------------------------ 1307

Qy     363 ProValAspAlaSerValGlyLeuAlaTrpLeuAlaValGlySerMetCysLeuPheIle 382
                 |||   :::   |||   |||   :::   |||      :::|||   ::::::
Db    1308 ------GACATCGCCAAGGGCTACGCCGCGGTCGTCGTGGTCTTCATCTGCGCCTACGTC 1361

Qy     383 AlaGlyPheAlaValGlyTrpGlyProIleProTrpLeuLeuMetSerGluIlePhePro 402
           ||||||||||||      |||||||||:::   ||||||:::   |||||||||||||||
Db    1362 GCCGGCTTCGCCTGGTCGTGGGGGCCCCTGGGCTGGCTCGTGCCGTCCGAGATCTTCCCG 1421

Qy     403 LeuHisValLysGlyValAlaThrGlyIleCysValLeuThrAsnTrpLeuMetAlaPhe 422
           |||   ::::::               |||   |||      |||            |||
Db    1422 CTGGAGATCCGCCCGGCGGGGCAGAGCATCAACGTCTCCGTCAACATGTTCTTCACCTTC 1481

Qy     423 LeuValThrLysGluPheSerSerLeuMetGluValLeuArgProTyrGlyAlaPheTrp 442
              :::   :::   |||   :::::::::         :::   :::|||   |||:::
Db    1482 TGCATCGCGCAGGCCTTCCTCACCATGCTCTGCCACTTCAAG---TTCGGCCTCTTCTAC 1538

Qy     443 LeuAlaSerAlaPheCysIlePheSerValLeuPheThrLeuPheCysValProGluThr 462
                 :::   :::   :::         :::|||      |||   :::|||||||||
Db    1539 TTCTTCGCCGGCTGGGTCGTCATCATGACCGTCTTTATCGCCTTCTTCCTGCCCGAGACC 1598

Qy     463 LysGlyLysThrLeuGluGlnIle 470
           |||         :::|||::::::
Db    1599 AAGAACGTGCCCATCGAGGAGATG 1622

RESULT 6
US-09-679-686B-15
; Sequence 15, Application US/09679686B
; Patent No. 6624343
; GENERAL INFORMATION:
;  APPLICANT: Allen, Stephen M.
;  APPLICANT:  Lightner, Jonathan E.
;  APPLICANT:  Rafalski, J. Antoni
;  APPLICANT:  Thorpe, Catherine J.
;  TITLE OF INVENTION: HEXOSE CARRIER PROTEINS
;  FILE REFERENCE: BB1160 US NA
;  CURRENT APPLICATION NUMBER: US/09/679,686B
;  CURRENT FILING DATE:  2003-01-16
;  PRIOR APPLICATION NUMBER: 60/081,131
;  PRIOR FILING DATE: 1998-04-09
;  PRIOR APPLICATION NUMBER: PCT/US99/07561
;  PRIOR FILING DATE: 1999-04-07
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 15
;   LENGTH: 1960
;   TYPE: DNA
;   ORGANISM: Glycine max
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (10)
;   OTHER INFORMATION: n = a, c, g or t
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (28)
;   OTHER INFORMATION: n = a, c, g or t
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (514)
;   OTHER INFORMATION: n = a, c, g or t
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (533)
;   OTHER INFORMATION: n = a, c, g or t
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (542)
;   OTHER INFORMATION: n = a, c, g or t
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (552)
;   OTHER INFORMATION: n = a, c, g or t
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (620)
;   OTHER INFORMATION: n = a, c, g or t
US-09-679-686B-15
```

```
Alignment Scores:
Pred. No.:              9.02e-44      Length:       1960
Score:                  518.00        Matches:      148
Percent Similarity:     47.55%        Conservative: 95
Best Local Similarity:  28.96%        Mismatches:   198
Query Match:            21.08%        Indels:       70
DB:                     4             Gaps:         16

US-09-886-954A-1 (1-477) x US-09-679-686B-15 (1-1960)

Qy          7 GluGluThrGlnProLeuLeuGlyPro---------------------------ProGly 17
              |||:::      |||   :::|||                              ||||||
Db        141 GAGAAAAAAATGCCTGCGGTAGGAGGTATTAGCAACGGAGGGGGCAAGGAGTACCCTGGA 200

Qy         18 GlySerAlaProArgGlyArgArgValPheLeuAlaAlaPheAlaAlaAlaLeuGlyPro 37
                       |||            |||   :::            ||||||:::|||
Db        201 AGCCTCACTCCT---------TTTGTGACGGTAACATGTATAGTTGCAGCCATGGGTGGG 251

Qy         38 LeuSerPheGlyPheAlaLeuGlyTyrSerSerProAlaIle---------ProSerLeu 54
              |||   ||||||:::   :::|||   |||                     |||   |||
Db        252 TTAATCTTCGGTTACGATATCGGAATTTCAGGTGGGGTGACATCCATGGATCCGTTTCTG 311

Qy         55 GlnArgAlaAlaProProAlaProArg--------------------------------- 63
                 :::      |||         |||
Db        312 CTCAAGTTTTTCCCGTCGGTGTTCCGGAAGAAGAATTCCGACAAAACGGTGAACCAGTAC 371

Qy         64 ------LeuAspAspAlaAlaAlaSerTrpPheGlyAlaValValThrLeuGlyAlaAla 81
                       |||            :::   |||   :::   :::   |||   |||
Db        372 TGTCAATACGACAGTCAGACACTGACGATGTTCACGTCGTCGCTGTATCTCGCCGCGTTG 431

Qy         82 AlaGlyGlyValLeuGlyGlyTrpLeuValAspArgAlaGlyArgLysLeuSerLeuLeu 101
                       ::::::         :::      |||   |||||||||   |||:::|||
Db        432 CTGTCGTCGTTGGTTGCCTCCACCGTCACACGTAGGTTCGGCCGRAAAMTCTCCATGCTT 491

Qy        102 LeuCysSerValProPheValAlaGlyPheAlaValIleThrAlaAlaGlnAspValTrp 121
                       :::   |||:::   |||      :::         |||      ||||||
Db        492 TTCGGAGGCTTGCTTTTCCTCGNCGGTGCCCTTATCAACGGNTTTGCCCANCACGTGTGG 551

Qy        122 MetLeuLeuGlyGlyArgLeuLeuThrGlyLeuAlaCysGlyValAlaSerLeuValAla 141
                 |||:::   ||||||:::|||   |||               |||:::
Db        552 NTGCTCATCGTGGGTCGGATCTTGCTCGGGTTCGGTATCCGGTTTGCCAATCAGTCTGTG 611

Qy        142 ProValTyrIleSerGluIleAlaTyrProAlaValArgGlyLeuLeuGlySerCysVal 161
              |||:::   :::||||||:::|||            ||||||   |||
Db        612 CCACTCTANCTATCTGAAATGGCTCCATACAAATATAGAGGAGCATTGAACATTGGCTTT 671

Qy        162 GlnLeuMetValValValGlyIleLeuLeuAlaTyrLeuAlaGlyTrpValLeuGlu--- 180
              ||||||   :::   ||||||||||||:::|||   :::      :::
Db        672 CAGTTGTCCATCACTGTTGGTATCCTTGTGGCCAATGTGTTGAACTATTTCTTTGCTAAA 731

Qy        181 ------------TrpArgTrp---LeuAlaValLeuGlyCysValProProSerLeuMet 195
                          |||   |||   |||::::::   |||   :::   |||:::|||:::
Db        732 ATCAAAGGTGGTTGGGGATGGAGGTTGAGTTTGGGAGGTGCTATGGTCCCTGCCCTTATA 791

Qy        196 LeuLeuLeu---MetCysPheMetProGluThrProArgPheLeuLeuThrGlnHisArg 214
              :::   :::            :::|||:::||||||      ::::::   :::
Db        792 ATCACAGTAGGATCACTAGTCCTTCCAGACACTCCCAATTCCATGATTGAAAGGGGTGAT 851

Qy        215 ArgGlnGluAlaMetAlaAlaLeuArgPheLeuTrpGly---------SerGluGlnGly 231
              |||::::::|||   |||   |||:::   :::   |||            ::::::
Db        852 CGCGAGAAGGCCAAGGCTCAGCTTCAGAGAATTCGCGGCATCGACAATGTTGATGAAGAG 911

Qy        232 TrpGluAspProProIleGlyAlaGluGlnSerPheHisLeuAla--------------- 246
              :::   |||            :::|||   |||      :::
Db        912 TTCAATGACCTTGTGGCAGCAAGTGAATCCTCTAGCCAAGTGGAGCACCCTTGGAGGAAC 971

Qy        247 LeuLeuArgGlnProGlyIleTyrLysProPheIle---IleGlyValSerLeuMetAla 265
              ||||||::::::         |||:::|||   :::   :::   |||   :::
Db        972 TTGTTGCAAAGAAAG------TACAGACCCCACCTCACCATGGCAGTGTTGATTCCATTC 1025

Qy        266 PheGlnGlnLeuSerGlyValAsnAlaValMetPheTyrAlaGluThrIlePheGluGlu 285
              ||||||||||||:::|||:::|||   :::||||||||||||      :::|||
Db       1026 TTCCAGCAACTCACTGGAATCAATGTCATCATGTTTTACGCGCCGGTCCTGTTCAGCTCC 1085

Qy        286 AlaLysPheLysAspSerSer------LeuAlaSerValValValGlyValIleGlnVal 303
                    |||||||||   :::      :::::::::|||:::   ||||||:::   |||
Db       1086 ATCGGGTTTAAGGATGATGCTGCTCTAATGTCAGCTGTGATCACCGGCGTTGTTAATGTT 1145

Qy        304 LeuPheThrAlaValAlaAlaLeuIleMetAspArgAlaGlyArgArgLeuLeuLeuVal 323
              :::   |||   |||:::         :::|||:::   |||||||||   |||   :::
Db       1146 GTCGCAACTTGTGTCTCAATTTATGGTGTTGACAAGTGGGGTAGGAGAGCCCTTTTCCTT 1205

Qy        324 LeuSerGlyValValMetValPheSer------------ThrSerAlaPheGlyAlaTyr 339
                    ||||||   |||:::                     :::|||   ||||||
Db       1206 GAAGGTGGAGTCCAAATGCTCATTTGCCAGGCTGTAGTTGCAGCTGCAATTGGAGCAAAG 1265

Qy        340 PheLysLeuThrGlnGlyGlyProGlyAsnSerSerHisValAlaIleSerAlaProVal 359
              |||      |||   |||   ||||||
Db       1266 TTTGGA---ACTGATGGGAACCCAGGT--------------------------------- 1289

Qy        360 SerAlaGlnProValAspAlaSerValGlyLeuAlaTrpLeuAlaValGlySerMetCys 379
                             |||               |||   :::   |||      :::|||
Db       1290 ---------------GATTTGCCAAAGTGGTATGCAATTGTTGTGGTTCTCTTCATTTGC 1334

Qy        380 LeuPheIleAlaGlyPheAlaValGlyTrpGlyProIleProTrpLeuLeuMetSerGlu 399
              ::::::::::::   ||||||      |||||||||:::   ||||||:::   ||||||
Db       1335 ATTTATGTATCAGCATTTGCCTGGTCATGGGGTCCCCTAGGTTGGTTGGTGCCTAGTGAG 1394

Qy        400 IlePheProLeuHisValLysGlyValAlaThrGlyIleCysValLeuThrAsnTrpLeu 419
              ||||||||||||   ::::::      |||      |||   |||      |||   |||
Db       1395 ATCTTTCCCTTGGAGATTCGTTCAGCTGCTCAGAGTATCAATGTGTCGGTGAACATGCTT 1454

Qy        420 MetAlaPheLeuValThrLysGluPheSerSerLeuMetGluValLeuArgProTyrGly 439
                    ||||||:::   :::   |||   :::::::::      ::::::   :::|||
Db       1455 TTCACTTTCTTGATTGCACAAGTCTTCTTGACAATGCTTTGCCACATGAAG---TTCGGC 1511

Qy        440 AlaPheTrpLeuAlaSerAlaPheCysIlePheSerValLeuPheThrLeuPheCysVal 459
                 |||         :::   |||   :::            |||      |||   :::
Db       1512 TTGTTCCTCTTCTTTGCCTTCTTCGTGTTGATCATGACATTCTTCGTCTACTTCTTCTTG 1571

Qy        460 ProGluThrLysGlyLysThrLeuGluGlnIle 470
              |||||||||||||||      :::|||::::::
Db       1572 CCCGAAACAAAGGGCATTCCAATTGAAGAAATG 1604

RESULT 7
US-09-643-597-135
; Sequence 135, Application US/09643597
; Patent No. 6426072
; GENERAL INFORMATION:
;  APPLICANT:  Wang, Tongtong
;  APPLICANT:  Fan, Liqun
;  APPLICANT:  Kalos, Michael D.
;  APPLICANT:  Bangur, Chaitanya S.
;  APPLICANT:  Hosken, Nancy
;  APPLICANT:  Fanger, Gary R.
;  APPLICANT:  Li, Samuel X.
;  APPLICANT:  Wang, Aijun
;  APPLICANT:  Skeiky, Yasir A.W.
;  APPLICANT:  Henderson, Robert A.
;  APPLICANT:  McNeill, Patricia D.
;  TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE:  210121.455C11
;  CURRENT APPLICATION NUMBER: US/09/643,597
;  CURRENT FILING DATE:  2000-08-21
;  NUMBER OF SEQ ID NOS:  369
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 135
;   LENGTH: 2856
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-09-643-597-135

Alignment Scores:
Pred. No.:              3.76e-43      Length:       2856
Score:                  514.50        Matches:      145
```

```
Percent Similarity:      46.92%        Conservative:    91
Best Local Similarity:   28.83%        Mismatches:      168
Query Match:             20.94%        Indels:          99
DB:                      4             Gaps:            14

US-09-886-954A-1 (1-477) x US-09-643-597-135 (1-2856)

Qy      25 ArgValPheLeuAlaAlaPheAlaAlaAlaLeuGlyProLeuSerPheGlyPheAlaLeu 44
           |||:::   ||||||         |||   ||||||   |||   ||||||:::
Db     210 CGCCTCATGCTGGCTGTGGGAGGAGCAGTGCTTGGCTCCCTGCAGTTTGGCTACAACACT 269

Qy      45 Gly--------------------------------------------------------- 45
           |||
Db     270 GGAGTCATCAATGCCCCCCAGAAGGTGATCGAGGAGTTCTACAACCAGACATGGGTCCAC 329

Qy      46 ---TyrSerSerProAlaIleProSerLeuGlnArgAlaAlaProProAlaProArgLeu 64
              |||            :::|||:::
Db     330 CGCTATGGGGAGAGCATCCTGCCCACC--------------------------------- 356

Qy      65 AspAspAlaAlaAlaSerTrpPheGlyAlaValValThrLeuGlyAlaAlaAlaGlyGly 84
                             |||      :::|||               :::   ||||||
Db     357 ---ACGCTCACCACGCTCTGGTCCCTCTCAGTGGCCATCTTT------TCTGTTGGGGGC 407

Qy      85 ValLeuGly------------GlyTrpLeuValAspArgAlaGlyArgLysLeuSerLeu 100
           ::::::|||            |||      |||:::|||   ||||||:::   |||:::
Db     408 ATGATTGGCTCCTTCTCTGTGGGCCTTTTCGTTAACCGCTTTGGCCGGCGGAATTCAATG 467

Qy     101 LeuLeuCysSer---ValProPheValAla------------GlyPheAlaValIleThr 115
           |||:::   :::   :::   ||||||:::            ||||||:::   :::
Db     468 CTGATGATGAACCTGCTGGCCTTCGTGTCCGCCGTGCTCATGGGCTTCTCGAAACTG--- 524

Qy     116 AlaAlaGlnAspValTrpMetLeuLeuGlyGlyArgLeuLeuThrGlyLeuAlaCysGly 135
                 :::         ||||||:::   ||||||   :::   |||:::   ||||||
Db     525 ---GGCAAGTCCTTTGAGATGCTGATCCTGGGCCGCTTCATCATCGGTGTGTACTGCGGC 581

Qy     136 ValAlaSerLeuValAlaProValTyrIleSerGluIleAlaTyrProAlaValArgGly 155
           :::   :::         |||:::|||:::   |||::::::      |||   ||||||
Db     582 CTGACCACAGGCTTCGTGCCCATGTATGTGGGTGAAGTGTCACCCACAGCCTTTCGTGGG 641

Qy     156 LeuLeuGlySerCysValGlnLeuMetValValValGlyIleLeuLeuAlaTyrLeuAla 175
              ||||||:::      ||||||   :::|||||||||||||||:::|||   :::
Db     642 GCCCTGGGCACCCTGCACCAGCTGGGCATCGTCGTCGGCATCCTCATCGCCCAGGTGTTC 701

Qy     176 GlyTrpValLeuGlu------------------------TrpArgTrpLeuAlaValLeu 187
           |||      |||:::                        |||      |||      :::
Db     702 GGC------CTGGACTCCATCATGGGCAACAAGGACCTGTGGCCCCTGCTGCTGAGCATC 755

Qy     188 GlyCysValProProSerLeuMetLeuLeuLeuMetCysPheMetProGluThrProArg 207
                 :::|||      |||      :::::::::   |||   ||||||:::||||||
Db     756 ATCTTCATCCCGGCCCTGCTGCAGTGCATCGTGCTGCCCTTCTGCCCCGAGAGTCCCCGC 815

Qy     208 PheLeuLeu---ThrGlnHisArgArgGlnGluAlaMetAlaAlaLeuArgPheLeuTrp 226
           |||||||||      ::::::            |||   :::   |||:::   |||
Db     816 TTCCTGCTCATCAACCGCAACGAGGAGAACCGGGCCAAGAGTGTGCTAAAGAAGCTGCGC 875

Qy     227 GlySerGluGlnGlyTrpGluAspProProIleGlyAlaGluGlnSerPheHis------ 244
           |||:::               |||               |||:::|||
Db     876 GGGACAGCTGACGTGACCCATGACCTGCAGGAGATGAAGGAAGAGAGTCGGCAGATGATG 935

Qy     245 ---------------------LeuAlaLeuLeuArgGlnProGlyIleTyrLysProPhe 257
                                |||   |||   |||   |||         :::|||
Db     936 CGGGAGAAGAAGGTCACCATCCTGGAGCTGTTCCGCTCCCCCGCCTACCGCCAGCCCATC 995

Qy     258 IleIleGlyValSerLeuMetAlaPheGlnGlnLeuSerGlyValAsnAlaValMetPhe 277
           :::|||   |||   |||         |||||||||||||||:::|||||||||   :::
Db     996 CTCATCGCTGTGGTGCTGCAGCTGTCCCAGCAGCTGTCTGGCATCAACGCTGTCTTCTAT 1055

Qy     278 TyrAlaGluThrIlePheGluGluAlaLysPheLysAspSerSerLeuAlaSerValVal 297
           |||:::   :::|||||||||:::|||      :::            |||::::::
Db    1056 TACTCCACGAGCATCTTCGAGAAGGCGGGGGTGCAGCAGCCTGTGTATGCCACCATTGGC 1115

Qy     298 ValGlyValIleGlnValLeuPheThrAlaValAlaAlaLeuIleMetAspArgAlaGly 317
              |||::::::         ||||||   |||:::      :::::::::|||||||||
```

```
Db    1116 TCCGGTATCGTCAACACGGCCTTCACTGTCGTGTCGCTGTTTGTGGTGGAGCGAGCAGGC 1175

Qy     318 ArgArgLeuLeuLeuValLeuSerGlyValValMetValPheSerThrSerAlaPheGly 337
           ||||||   |||   ::::::
Db    1176 CGGCGGACCCTGCACCTCATA--------------------------------------- 1196

Qy     338 AlaTyrPheLysLeuThrGlnGlyGlyProGlyAsnSerSerHisValAlaIleSerAla 357
                                   |||            :::      ||||||
Db    1197 ------------------------GGCCTCGCTGGCATGGCGGGTTGTGCCATACTCATG 1232

Qy     358 ProValSerAlaGlnProValAspAlaSerValGlyLeuAlaTrpLeuAlaValGlySer 377
              ::::::         ::::::            :::::::::|||::::::   :::
Db    1233 ACCATCGCGCTAGCACTGCTGGAGCAGCTACCCTGGATGTCCTATCTGAGCATCGTGGCC 1292

Qy     378 MetCysLeuPheIleAlaGlyPheAlaValGlyTrpGlyProIleProTrpLeuLeuMet 397
           :::      |||:::|||   |||   ||||||   |||||||||||||||   ::::::
Db    1293 ATCTTTGGCTTTGTGGCCTTCTTTGAAGTGGGTCCTGGCCCCATCCCATGGTTCATCGTG 1352

Qy     398 SerGluIlePheProLeuHisValLysGlyValAlaThrGlyIleCysValLeuThrAsn 417
           :::|||:::|||            :::      |||      :::         :::|||
Db    1353 GCTGAACTCTTCAGCCAGGGTCCACGTCCAGCTGCCATTGCCGTTGCAGGCTTCTCCAAC 1412

Qy     418 TrpLeuMetAlaPheLeuValThrLysGluPheSerSerLeuMetGluValLeuArgPro 437
           |||         |||:::|||         |||      :::   ::::::      |||
Db    1413 TGGACCTCAAATTTCATTGTGGGCATGTGCTTCCAGTATGTGGAGCAACTGTGTGGTCCC 1472

Qy     438 TyrGlyAlaPheTrpLeuAlaSerAlaPheCysIlePheSerValLeu------------ 453
           |||                        |||   ||||||:::||||||
Db    1473 TAC---------------------GTCTTCATCATCTTCACTGTGCTCCTGGTTCTGTTC 1511

Qy     454 ------PheThrLeuPheCysValProGluThrLysGlyLysThrLeuGluGlnIleThr 471
                 ||||||   |||   ||||||||||||||||||:::|||   ::::::|||
Db    1512 TTCATCTTCACCTACTTCAAAGTTCCTGAGACTAAAGGCCGGACCTTCGATGAGATCGCT 1571

Qy     472 AlaHisPhe 474
           :::   |||
Db    1572 TCCGGCTTC 1580

RESULT 8
US-09-480-884A-135
; Sequence 135, Application US/09480884A
; Patent No. 6482597
; GENERAL INFORMATION:
;  APPLICANT: Wang, Tongtong
;  APPLICANT:  Fan, Liqun
;  APPLICANT:  Hosken, Nancy A.
;  APPLICANT:  Kalos, Michael D.
;  APPLICANT:  Fanger, Gary R.
;  TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE: 210121.455C6
;  CURRENT APPLICATION NUMBER: US/09/480,884A
;  CURRENT FILING DATE:  2001-08-27
;  NUMBER OF SEQ ID NOS: 330
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 135
;   LENGTH: 2856
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-09-480-884A-135

Alignment Scores:
Pred. No.:              3.76e-43        Length:         2856
Score:                  514.50          Matches:        145
Percent Similarity:     46.92%          Conservative:   91
Best Local Similarity:  28.83%          Mismatches:     168
Query Match:            20.94%          Indels:         99
DB:                     4               Gaps:           14

US-09-886-954A-1 (1-477) x US-09-480-884A-135 (1-2856)

Qy      25 ArgValPheLeuAlaAlaPheAlaAlaAlaLeuGlyProLeuSerPheGlyPheAlaLeu 44
           |||:::   ||||||         |||   ||||||   |||   ||||||:::
```

```
Db      210 CGCCTCATGCTGGCTGTGGGAGGAGCAGTGCTTGGCTCCCTGCAGTTTGGCTACAACACT 269

Qy       45 Gly--------------------------------------------------------- 45
            |||
Db      270 GGAGTCATCAATGCCCCCCAGAAGGTGATCGAGGAGTTCTACAACCAGACATGGGTCCAC 329

Qy       46 ---TyrSerSerProAlaIleProSerLeuGlnArgAlaAlaProProAlaProArgLeu 64
               |||            :::|||:::
Db      330 CGCTATGGGGAGAGCATCCTGCCCACC--------------------------------- 356

Qy       65 AspAspAlaAlaAlaSerTrpPheGlyAlaValValThrLeuGlyAlaAlaAlaGlyGly 84
                              |||      :::|||               :::   ||||||
Db      357 ---ACGCTCACCACGCTCTGGTCCCTCTCAGTGGCCATCTTT------TCTGTTGGGGGC 407

Qy       85 ValLeuGly------------GlyTrpLeuValAspArgAlaGlyArgLysLeuSerLeu 100
            ::::::|||            |||      |||:::|||   ||||||:::   |||:::
Db      408 ATGATTGGCTCCTTCTCTGTGGGCCTTTTCGTTAACCGCTTTGGCCGGCGGAATTCAATG 467

Qy      101 LeuLeuCysSer---ValProPheValAla------------GlyPheAlaValIleThr 115
            |||:::   :::   :::   ||||||:::            ||||||:::   :::
Db      468 CTGATGATGAACCTGCTGGCCTTCGTGTCCGCCGTGCTCATGGGCTTCTCGAAACTG--- 524

Qy      116 AlaAlaGlnAspValTrpMetLeuLeuGlyGlyArgLeuLeuThrGlyLeuAlaCysGly 135
                  :::         ||||||:::   ||||||   :::   |||:::   ||||||
Db      525 ---GGCAAGTCCTTTGAGATGCTGATCCTGGGCCGCTTCATCATCGGTGTGTACTGCGGC 581

Qy      136 ValAlaSerLeuValAlaProValTyrIleSerGluIleAlaTyrProAlaValArgGly 155
            :::   :::         |||:::|||:::   |||::::::      |||   ||||||
Db      582 CTGACCACAGGCTTCGTGCCCATGTATGTGGGTGAAGTGTCACCCACAGCCTTTCGTGGG 641

Qy      156 LeuLeuGlySerCysValGlnLeuMetValValValGlyIleLeuLeuAlaTyrLeuAla 175
               ||||||:::      ||||||   :::|||||||||||||||:::|||   :::
Db      642 GCCCTGGGCACCCTGCACCAGCTGGGCATCGTCGTCGGCATCCTCATCGCCCAGGTGTTC 701

Qy      176 GlyTrpValLeuGlu------------------------TrpArgTrpLeuAlaValLeu 187
            |||      |||:::                        |||      |||      :::
Db      702 GGC------CTGGACTCCATCATGGGCAACAAGGACCTGTGGCCCCTGCTGCTGAGCATC 755

Qy      188 GlyCysValProProSerLeuMetLeuLeuLeuMetCysPheMetProGluThrProArg 207
                  :::|||      |||      :::::::::   |||   ||||||:::||||||
Db      756 ATCTTCATCCCGGCCCTGCTGCAGTGCATCGTGCTGCCCTTCTGCCCCGAGAGTCCCCGC 815

Qy      208 PheLeuLeu---ThrGlnHisArgArgGlnGluAlaMetAlaAlaLeuArgPheLeuTrp 226
            |||||||||      ::::::            |||   :::   |||:::   |||
Db      816 TTCCTGCTCATCAACCGCAACGAGGAGAACCGGGCCAAGAGTGTGCTAAAGAAGCTGCGC 875

Qy      227 GlySerGluGlnGlyTrpGluAspProProIleGlyAlaGluGlnSerPheHis------ 244
            |||:::               |||               |||:::|||
Db      876 GGGACAGCTGACGTGACCCATGACCTGCAGGAGATGAAGGAAGAGAGTCGGCAGATGATG 935

Qy      245 ---------------------LeuAlaLeuLeuArgGlnProGlyIleTyrLysProPhe 257
                                 |||   |||   |||   |||         :::|||
Db      936 CGGGAGAAGAAGGTCACCATCCTGGAGCTGTTCCGCTCCCCCGCCTACCGCCAGCCCATC 995

Qy      258 IleIleGlyValSerLeuMetAlaPheGlnGlnLeuSerGlyValAsnAlaValMetPhe 277
            :::|||   |||   |||         |||||||||||||||:::|||||||||   :::
Db      996 CTCATCGCTGTGGTGCTGCAGCTGTCCCAGCAGCTGTCTGGCATCAACGCTGTCTTCTAT 1055

Qy      278 TyrAlaGluThrIlePheGluGluAlaLysPheLysAspSerSerLeuAlaSerValVal 297
            |||:::   :::|||||||||:::|||      :::            |||::::::
Db     1056 TACTCCACGAGCATCTTCGAGAAGGCGGGGGTGCAGCAGCCTGTGTATGCCACCATTGGC 1115

Qy      298 ValGlyValIleGlnValLeuPheThrAlaValAlaAlaLeuIleMetAspArgAlaGly 317
               |||::::::         ||||||   |||:::      :::::::::|||||||||
Db     1116 TCCGGTATCGTCAACACGGCCTTCACTGTCGTGTCGCTGTTTGTGGTGGAGCGAGCAGGC 1175

Qy      318 ArgArgLeuLeuLeuValLeuSerGlyValValMetValPheSerThrSerAlaPheGly 337
            ||||||   |||   ::::::
Db     1176 CGGCGGACCCTGCACCTCATA--------------------------------------- 1196

Qy      338 AlaTyrPheLysLeuThrGlnGlyGlyProGlyAsnSerSerHisValAlaIleSerAla 357
                                    |||            :::      ||||||
Db     1197 ------------------------GGCCTCGCTGGCATGGCGGGTTGTGCCATACTCATG 1232

Qy      358 ProValSerAlaGlnProValAspAlaSerValGlyLeuAlaTrpLeuAlaValGlySer 377
               ::::::         ::::::            :::::::::|||::::::   :::
Db     1233 ACCATCGCGCTAGCACTGCTGGAGCAGCTACCCTGGATGTCCTATCTGAGCATCGTGGCC 1292

Qy      378 MetCysLeuPheIleAlaGlyPheAlaValGlyTrpGlyProIleProTrpLeuLeuMet 397
            :::      |||:::|||   |||   ||||||   |||||||||||||||   ::::::
Db     1293 ATCTTTGGCTTTGTGGCCTTCTTTGAAGTGGGTCCTGGCCCCATCCCATGGTTCATCGTG 1352

Qy      398 SerGluIlePheProLeuHisValLysGlyValAlaThrGlyIleCysValLeuThrAsn 417
            :::|||:::|||            :::      |||      :::         :::|||
Db     1353 GCTGAACTCTTCAGCCAGGGTCCACGTCCAGCTGCCATTGCCGTTGCAGGCTTCTCCAAC 1412

Qy      418 TrpLeuMetAlaPheLeuValThrLysGluPheSerSerLeuMetGluValLeuArgPro 437
            |||         |||:::|||         |||      :::   ::::::      |||
Db     1413 TGGACCTCAAATTTCATTGTGGGCATGTGCTTCCAGTATGTGGAGCAACTGTGTGGTCCC 1472

Qy      438 TyrGlyAlaPheTrpLeuAlaSerAlaPheCysIlePheSerValLeu------------ 453
            |||                        |||   ||||||:::||||||
Db     1473 TAC---------------------GTCTTCATCATCTTCACTGTGCTCCTGGTTCTGTTC 1511

Qy      454 ------PheThrLeuPheCysValProGluThrLysGlyLysThrLeuGluGlnIleThr 471
                  ||||||   |||   ||||||||||||||||||:::|||   ::::::|||
Db     1512 TTCATCTTCACCTACTTCAAAGTTCCTGAGACTAAAGGCCGGACCTTCGATGAGATCGCT 1571

Qy      472 AlaHisPhe 474
            :::   |||
Db     1572 TCCGGCTTC 1580
```

```
RESULT 9
US-09-542-615A-135
; Sequence 135, Application US/09542615A
; Patent No. 6518256
; GENERAL INFORMATION:
;  APPLICANT:  Wang, Tongtong
;  APPLICANT:  Fan, Liqun
;  APPLICANT:  Kalos, Michael D.
;  APPLICANT:  Bangur, Chaitanya S.
;  APPLICANT:  Hosken, Nancy A.
;  APPLICANT:  Fanger, Gary R.
;  TITLE OF INVENTION:  COMPOUNDS AND METHODS FOR THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE:  210121.455C8
;  CURRENT APPLICATION NUMBER: US/09/542,615A
;  CURRENT FILING DATE:  2000-04-14
;  NUMBER OF SEQ ID NOS:  350
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 135
;   LENGTH: 2856
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-09-542-615A-135

Alignment Scores:
Pred. No.:               3.76e-43        Length:        2856
Score:                   514.50          Matches:       145
Percent Similarity:      46.92%          Conservative:  91
Best Local Similarity:   28.83%          Mismatches:    168
Query Match:             20.94%          Indels:        99
DB:                      4               Gaps:          14

US-09-886-954A-1 (1-477) x US-09-542-615A-135 (1-2856)

Qy       25 ArgValPheLeuAlaAlaPheAlaAlaAlaLeuGlyProLeuSerPheGlyPheAlaLeu 44
            |||:::   ||||||         |||   ||||||   |||   ||||||:::
Db      210 CGCCTCATGCTGGCTGTGGGAGGAGCAGTGCTTGGCTCCCTGCAGTTTGGCTACAACACT 269

Qy       45 Gly--------------------------------------------------------- 45
            |||
Db      270 GGAGTCATCAATGCCCCCCAGAAGGTGATCGAGGAGTTCTACAACCAGACATGGGTCCAC 329

Qy       46 ---TyrSerSerProAlaIleProSerLeuGlnArgAlaAlaProProAlaProArgLeu 64
               |||            :::|||:::
```

```
Db        330 CGCTATGGGGAGAGCATCCTGCCCACC--------------------------------- 356

Qy         65 AspAspAlaAlaAlaSerTrpPheGlyAlaValValThrLeuGlyAlaAlaAlaGlyGly 84
                                |||      :::|||               :::   ||||||
Db        357 ---ACGCTCACCACGCTCTGGTCCCTCTCAGTGGCCATCTTT------TCTGTTGGGGGC 407

Qy         85 ValLeuGly------------GlyTrpLeuValAspArgAlaGlyArgLysLeuSerLeu 100
              ::::::|||            |||      |||:::|||   ||||||:::   |||:::
Db        408 ATGATTGGCTCCTTCTCTGTGGGCCTTTTCGTTAACCGCTTTGGCCGGCGGAATTCAATG 467

Qy        101 LeuLeuCysSer---ValProPheValAla------------GlyPheAlaValIleThr 115
              |||:::   :::   :::   ||||||:::            ||||||:::   :::
Db        468 CTGATGATGAACCTGCTGGCCTTCGTGTCCGCCGTGCTCATGGGCTTCTCGAAACTG--- 524

Qy        116 AlaAlaGlnAspValTrpMetLeuLeuGlyGlyArgLeuLeuThrGlyLeuAlaCysGly 135
                    :::         ||||||:::   ||||||   :::   |||:::   ||||||
Db        525 ---GGCAAGTCCTTTGAGATGCTGATCCTGGGCCGCTTCATCATCGGTGTGTACTGCGGC 581

Qy        136 ValAlaSerLeuValAlaProValTyrIleSerGluIleAlaTyrProAlaValArgGly 155
              :::   :::         |||:::|||:::   |||::::::      |||   ||||||
Db        582 CTGACCACAGGCTTCGTGCCCATGTATGTGGGTGAAGTGTCACCCACAGCCTTTCGTGGG 641

Qy        156 LeuLeuGlySerCysValGlnLeuMetValValValGlyIleLeuLeuAlaTyrLeuAla 175
                 ||||||:::      ||||||   :::|||||||||||||||:::|||   :::
Db        642 GCCCTGGGCACCCTGCACCAGCTGGGCATCGTCGTCGGCATCCTCATCGCCCAGGTGTTC 701

Qy        176 GlyTrpValLeuGlu------------------------TrpArgTrpLeuAlaValLeu 187
              |||      |||:::                        |||      |||      :::
Db        702 GGC------CTGGACTCCATCATGGGCAACAAGGACCTGTGGCCCCTGCTGCTGAGCATC 755

Qy        188 GlyCysValProProSerLeuMetLeuLeuLeuMetCysPheMetProGluThrProArg 207
                    :::|||      |||      :::::::::   |||   ||||||:::||||||
Db        756 ATCTTCATCCCGGCCCTGCTGCAGTGCATCGTGCTGCCCTTCTGCCCCGAGAGTCCCCGC 815

Qy        208 PheLeuLeu---ThrGlnHisArgArgGlnGluAlaMetAlaAlaLeuArgPheLeuTrp 226
              |||||||||      ::::::            |||   :::   |||:::   |||
Db        816 TTCCTGCTCATCAACCGCAACGAGGAGAACCGGGCCAAGAGTGTGCTAAAGAAGCTGCGC 875

Qy        227 GlySerGluGlnGlyTrpGluAspProProIleGlyAlaGluGlnSerPheHis------ 244
              |||:::               |||               |||:::|||
Db        876 GGGACAGCTGACGTGACCCATGACCTGCAGGAGATGAAGGAAGAGAGTCGGCAGATGATG 935

Qy        245 ---------------------LeuAlaLeuLeuArgGlnProGlyIleTyrLysProPhe 257
                                   |||   |||   |||   |||         :::|||
Db        936 CGGGAGAAGAAGGTCACCATCCTGGAGCTGTTCCGCTCCCCCGCCTACCGCCAGCCCATC 995

Qy        258 IleIleGlyValSerLeuMetAlaPheGlnGlnLeuSerGlyValAsnAlaValMetPhe 277
              :::|||   |||   |||         |||||||||||||||:::|||||||||   :::
Db        996 CTCATCGCTGTGGTGCTGCAGCTGTCCCAGCAGCTGTCTGGCATCAACGCTGTCTTCTAT 1055

Qy        278 TyrAlaGluThrIlePheGluGluAlaLysPheLysAspSerSerLeuAlaSerValVal 297
              |||:::   :::|||||||||:::|||      :::            |||::::::
Db       1056 TACTCCACGAGCATCTTCGAGAAGGCGGGGGTGCAGCAGCCTGTGTATGCCACCATTGGC 1115

Qy        298 ValGlyValIleGlnValLeuPheThrAlaValAlaAlaLeuIleMetAspArgAlaGly 317
                 |||::::::         ||||||   |||:::      :::::::::|||||||||
Db       1116 TCCGGTATCGTCAACACGGCCTTCACTGTCGTGTCGCTGTTTGTGGTGGAGCGAGCAGGC 1175

Qy        318 ArgArgLeuLeuLeuValLeuSerGlyValValMetValPheSerThrSerAlaPheGly 337
              ||||||   |||   ::::::
Db       1176 CGGCGGACCCTGCACCTCATA--------------------------------------- 1196

Qy        338 AlaTyrPheLysLeuThrGlnGlyGlyProGlyAsnSerSerHisValAlaIleSerAla 357
                                      |||            :::      ||||||
Db       1197 ------------------------GGCCTCGCTGGCATGGCGGGTTGTGCCATACTCATG 1232

Qy        358 ProValSerAlaGlnProValAspAlaSerValGlyLeuAlaTrpLeuAlaValGlySer 377
                 ::::::         ::::::            :::::::::|||::::::   :::
Db       1233 ACCATCGCGCTAGCACTGCTGGAGCAGCTACCCTGGATGTCCTATCTGAGCATCGTGGCC 1292

Qy        378 MetCysLeuPheIleAlaGlyPheAlaValGlyTrpGlyProIleProTrpLeuLeuMet 397
              :::      |||:::|||   |||   ||||||   |||||||||||||||   ::::::
Db       1293 ATCTTTGGCTTTGTGGCCTTCTTTGAAGTGGGTCCTGGCCCCATCCCATGGTTCATCGTG 1352

Qy        398 SerGluIlePheProLeuHisValLysGlyValAlaThrGlyIleCysValLeuThrAsn 417
              :::|||:::|||            :::      |||      :::         :::|||
Db       1353 GCTGAACTCTTCAGCCAGGGTCCACGTCCAGCTGCCATTGCCGTTGCAGGCTTCTCCAAC 1412

Qy        418 TrpLeuMetAlaPheLeuValThrLysGluPheSerSerLeuMetGluValLeuArgPro 437
              |||         |||:::|||         |||      :::   ::::::      |||
Db       1413 TGGACCTCAAATTTCATTGTGGGCATGTGCTTCCAGTATGTGGAGCAACTGTGTGGTCCC 1472

Qy        438 TyrGlyAlaPheTrpLeuAlaSerAlaPheCysIlePheSerValLeu------------ 453
              |||                        |||   ||||||:::||||||
Db       1473 TAC---------------------GTCTTCATCATCTTCACTGTGCTCCTGGTTCTGTTC 1511

Qy        454 ------PheThrLeuPheCysValProGluThrLysGlyLysThrLeuGluGlnIleThr 471
                    ||||||   |||   ||||||||||||||||||:::|||   ::::::|||
Db       1512 TTCATCTTCACCTACTTCAAAGTTCCTGAGACTAAAGGCCGGACCTTCGATGAGATCGCT 1571

Qy        472 AlaHisPhe 474
              :::   |||
Db       1572 TCCGGCTTC 1580

RESULT 10
US-09-606-421B-135
; Sequence 135, Application US/09606421B
; Patent No. 6531315
; GENERAL INFORMATION:
;  APPLICANT: Wang, Tongtong
;  APPLICANT:  Fan, Liqun
;  APPLICANT:  Kalos, Michael D.
;  APPLICANT:  Bangur, Chaitanya S.
;  APPLICANT:  Hosken, Nancy
;  APPLICANT:  Fanger, Gary R.
;  APPLICANT:  Li, Samuel X.
;  APPLICANT:  Wang, Aijun
;  APPLICANT:  Skeiky, Yasir A.W.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE: 210121.455C9
;  CURRENT APPLICATION NUMBER: US/09/606,421B
;  CURRENT FILING DATE:  2000-06-28
;  NUMBER OF SEQ ID NOS: 358
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 135
;   LENGTH: 2856
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-09-606-421B-135

Alignment Scores:
Pred. No.:               3.76e-43      Length:          2856
Score:                   514.50        Matches:         145
Percent Similarity:      46.92%        Conservative:    91
Best Local Similarity:   28.83%        Mismatches:      168
Query Match:             20.94%        Indels:          99
DB:                      4             Gaps:            14

US-09-886-954A-1 (1-477) x US-09-606-421B-135 (1-2856)

Qy         25 ArgValPheLeuAlaAlaPheAlaAlaAlaLeuGlyProLeuSerPheGlyPheAlaLeu 44
              |||:::   ||||||         |||   ||||||   |||   ||||||:::
Db        210 CGCCTCATGCTGGCTGTGGGAGGAGCAGTGCTTGGCTCCCTGCAGTTTGGCTACAACACT 269

Qy         45 Gly--------------------------------------------------------- 45
              |||
Db        270 GGAGTCATCAATGCCCCCCAGAAGGTGATCGAGGAGTTCTACAACCAGACATGGGTCCAC 329

Qy         46 ---TyrSerSerProAlaIleProSerLeuGlnArgAlaAlaProProAlaProArgLeu 64
                 |||            :::|||:::
Db        330 CGCTATGGGGAGAGCATCCTGCCCACC--------------------------------- 356

Qy         65 AspAspAlaAlaAlaSerTrpPheGlyAlaValValThrLeuGlyAlaAlaAlaGlyGly 84
                                |||      :::|||               :::   ||||||
Db        357 ---ACGCTCACCACGCTCTGGTCCCTCTCAGTGGCCATCTTT------TCTGTTGGGGGC 407
```

```
Qy           85 ValLeuGly------------GlyTrpLeuValAspArgAlaGlyArgLysLeuSerLeu 100
                ::::::|||            |||      |||:::|||   ||||||:::   |||:::
Db          408 ATGATTGGCTCCTTCTCTGTGGGCCTTTTCGTTAACCGCTTTGGCCGGCGGAATTCAATG 467

Qy          101 LeuLeuCysSer---ValProPheValAla------------GlyPheAlaValIleThr 115
                |||:::   :::   :::   ||||||:::            ||||||:::   :::
Db          468 CTGATGATGAACCTGCTGGCCTTCGTGTCCGCCGTGCTCATGGGCTTCTCGAAACTG--- 524

Qy          116 AlaAlaGlnAspValTrpMetLeuLeuGlyGlyArgLeuLeuThrGlyLeuAlaCysGly 135
                      :::         ||||||:::   ||||||   :::   |||:::   ||||||
Db          525 ---GGCAAGTCCTTTGAGATGCTGATCCTGGGCCGCTTCATCATCGGTGTGTACTGCGGC 581

Qy          136 ValAlaSerLeuValAlaProValTyrIleSerGluIleAlaTyrProAlaValArgGly 155
                :::   :::         |||:::|||:::   |||::::::      |||   ||||||
Db          582 CTGACCACAGGCTTCGTGCCCATGTATGTGGGTGAAGTGTCACCCACAGCCTTTCGTGGG 641

Qy          156 LeuLeuGlySerCysValGlnLeuMetValValValGlyIleLeuLeuAlaTyrLeuAla 175
                   ||||||:::      ||||||   :::|||||||||||||||:::|||   :::
Db          642 GCCCTGGGCACCCTGCACCAGCTGGGCATCGTCGTCGGCATCCTCATCGCCCAGGTGTTC 701

Qy          176 GlyTrpValLeuGlu------------------------TrpArgTrpLeuAlaValLeu 187
                |||      |||:::                        |||      |||      :::
Db          702 GGC------CTGGACTCCATCATGGGCAACAAGGACCTGTGGCCCCTGCTGCTGAGCATC 755

Qy          188 GlyCysValProProSerLeuMetLeuLeuLeuMetCysPheMetProGluThrProArg 207
                      :::|||      |||      :::::::::   |||   ||||||:::||||||
Db          756 ATCTTCATCCCGGCCCTGCTGCAGTGCATCGTGCTGCCCTTCTGCCCCGAGAGTCCCCGC 815

Qy          208 PheLeuLeu---ThrGlnHisArgArgGlnGluAlaMetAlaAlaLeuArgPheLeuTrp 226
                |||||||||      ::::::            |||   :::   |||:::   |||
Db          816 TTCCTGCTCATCAACCGCAACGAGGAGAACCGGGCCAAGAGTGTGCTAAAGAAGCTGCGC 875

Qy          227 GlySerGluGlnGlyTrpGluAspProProIleGlyAlaGluGlnSerPheHis------ 244
                |||:::               |||               |||:::|||
Db          876 GGGACAGCTGACGTGACCCATGACCTGCAGGAGATGAAGGAAGAGAGTCGGCAGATGATG 935

Qy          245 ---------------------LeuAlaLeuLeuArgGlnProGlyIleTyrLysProPhe 257
                                     |||   |||   |||   |||         :::|||
Db          936 CGGGAGAAGAAGGTCACCATCCTGGAGCTGTTCCGCTCCCCCGCCTACCGCCAGCCCATC 995

Qy          258 IleIleGlyValSerLeuMetAlaPheGlnGlnLeuSerGlyValAsnAlaValMetPhe 277
                :::|||   |||   |||         |||||||||||||||:::|||||||||   :::
Db          996 CTCATCGCTGTGGTGCTGCAGCTGTCCCAGCAGCTGTCTGGCATCAACGCTGTCTTCTAT 1055

Qy          278 TyrAlaGluThrIlePheGluGluAlaLysPheLysAspSerSerLeuAlaSerValVal 297
                |||:::   :::|||||||||:::|||      :::            |||::::::
Db         1056 TACTCCACGAGCATCTTCGAGAAGGCGGGGGTGCAGCAGCCTGTGTATGCCACCATTGGC 1115

Qy          298 ValGlyValIleGlnValLeuPheThrAlaValAlaAlaLeuIleMetAspArgAlaGly 317
                   |||::::::         ||||||   |||:::      :::::::::|||||||||
Db         1116 TCCGGTATCGTCAACACGGCCTTCACTGTCGTGTCGCTGTTTGTGGTGGAGCGAGCAGGC 1175

Qy          318 ArgArgLeuLeuLeuValLeuSerGlyValValMetValPheSerThrSerAlaPheGly 337
                ||||||   |||   ::::::
Db         1176 CGGCGGACCCTGCACCTCATA--------------------------------------- 1196

Qy          338 AlaTyrPheLysLeuThrGlnGlyGlyProGlyAsnSerSerHisValAlaIleSerAla 357
                                        |||            :::      ||||||
Db         1197 ------------------------GGCCTCGCTGGCATGGCGGGTTGTGCCATACTCATG 1232

Qy          358 ProValSerAlaGlnProValAspAlaSerValGlyLeuAlaTrpLeuAlaValGlySer 377
                   ::::::         ::::::            :::::::::|||::::::   :::
Db         1233 ACCATCGCGCTAGCACTGCTGGAGCAGCTACCCTGGATGTCCTATCTGAGCATCGTGGCC 1292

Qy          378 MetCysLeuPheIleAlaGlyPheAlaValGlyTrpGlyProIleProTrpLeuLeuMet 397
                :::      |||:::|||   |||   ||||||   |||||||||||||||   ::::::
Db         1293 ATCTTTGGCTTTGTGGCCTTCTTTGAAGTGGGTCCTGGCCCCATCCCATGGTTCATCGTG 1352

Qy          398 SerGluIlePheProLeuHisValLysGlyValAlaThrGlyIleCysValLeuThrAsn 417
                :::|||:::|||            :::      |||      :::         :::|||
Db         1353 GCTGAACTCTTCAGCCAGGGTCCACGTCCAGCTGCCATTGCCGTTGCAGGCTTCTCCAAC 1412

Qy          418 TrpLeuMetAlaPheLeuValThrLysGluPheSerSerLeuMetGluValLeuArgPro 437
                |||         |||:::|||         |||      :::   ::::::      |||
Db         1413 TGGACCTCAAATTTCATTGTGGGCATGTGCTTCCAGTATGTGGAGCAACTGTGTGGTCCC 1472

Qy          438 TyrGlyAlaPheTrpLeuAlaSerAlaPheCysIlePheSerValLeu------------ 453
                |||                        |||   ||||||:::||||||
Db         1473 TAC---------------------GTCTTCATCATCTTCACTGTGCTCCTGGTTCTGTTC 1511

Qy          454 ------PheThrLeuPheCysValProGluThrLysGlyLysThrLeuGluGlnIleThr 471
                      ||||||   |||   ||||||||||||||||||:::|||   ::::::|||
Db         1512 TTCATCTTCACCTACTTCAAAGTTCCTGAGACTAAAGGCCGGACCTTCGATGAGATCGCT 1571

Qy          472 AlaHisPhe 474
                :::   |||
Db         1572 TCCGGCTTC 1580

RESULT 11
US-09-221-107-135
; Sequence 135, Application US/09221107
; Patent No. 6660838
; GENERAL INFORMATION:
;  APPLICANT: Wang, Tongtong
;  TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY OF LUNG CANCER
;  FILE REFERENCE: 210121.455C2
;  CURRENT APPLICATION NUMBER: US/09/221,107
;  CURRENT FILING DATE: 1998-12-22
;  NUMBER OF SEQ ID NOS: 161
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 135
;   LENGTH: 2856
;   TYPE: DNA
;   ORGANISM: Human
US-09-221-107-135

Alignment Scores:
Pred. No.:                3.76e-43        Length:         2856
Score:                    514.50          Matches:        145
Percent Similarity:       46.92%          Conservative:   91
Best Local Similarity:    28.83%          Mismatches:     168
Query Match:              20.94%          Indels:         99
DB:                       4               Gaps:           14

US-09-886-954A-1 (1-477) x US-09-221-107-135 (1-2856)

Qy           25 ArgValPheLeuAlaAlaPheAlaAlaAlaLeuGlyProLeuSerPheGlyPheAlaLeu 44
                |||:::   ||||||         |||   ||||||   |||   ||||||:::
Db          210 CGCCTCATGCTGGCTGTGGGAGGAGCAGTGCTTGGCTCCCTGCAGTTTGGCTACAACACT 269

Qy           45 Gly--------------------------------------------------------- 45
                |||
Db          270 GGAGTCATCAATGCCCCCCAGAAGGTGATCGAGGAGTTCTACAACCAGACATGGGTCCAC 329

Qy           46 ---TyrSerSerProAlaIleProSerLeuGlnArgAlaAlaProProAlaProArgLeu 64
                   |||            :::|||:::
Db          330 CGCTATGGGGAGAGCATCCTGCCCACC--------------------------------- 356

Qy           65 AspAspAlaAlaAlaSerTrpPheGlyAlaValValThrLeuGlyAlaAlaAlaGlyGly 84
                                  |||      :::|||               :::   ||||||
Db          357 ---ACGCTCACCACGCTCTGGTCCCTCTCAGTGGCCATCTTT------TCTGTTGGGGGC 407

Qy           85 ValLeuGly------------GlyTrpLeuValAspArgAlaGlyArgLysLeuSerLeu 100
                ::::::|||            |||      |||:::|||   ||||||:::   |||:::
Db          408 ATGATTGGCTCCTTCTCTGTGGGCCTTTTCGTTAACCGCTTTGGCCGGCGGAATTCAATG 467

Qy          101 LeuLeuCysSer---ValProPheValAla------------GlyPheAlaValIleThr 115
                |||:::   :::   :::   ||||||:::            ||||||:::   :::
Db          468 CTGATGATGAACCTGCTGGCCTTCGTGTCCGCCGTGCTCATGGGCTTCTCGAAACTG--- 524

Qy          116 AlaAlaGlnAspValTrpMetLeuLeuGlyGlyArgLeuLeuThrGlyLeuAlaCysGly 135
                      :::         ||||||:::   ||||||   :::   |||:::   ||||||
Db          525 ---GGCAAGTCCTTTGAGATGCTGATCCTGGGCCGCTTCATCATCGGTGTGTACTGCGGC 581

Qy          136 ValAlaSerLeuValAlaProValTyrIleSerGluIleAlaTyrProAlaValArgGly 155
```

```
             :::   :::         |||:::|||:::   |||::::::      |||   ||||||
Db       582 CTGACCACAGGCTTCGTGCCCATGTATGTGGGTGAAGTGTCACCCACAGCCTTTCGTGGG 641

Qy       156 LeuLeuGlySerCysValGlnLeuMetValValValGlyIleLeuLeuAlaTyrLeuAla 175
                ||||||:::      ||||||   :::|||||||||||||||:::|||   :::
Db       642 GCCCTGGGCACCCTGCACCAGCTGGGCATCGTCGTCGGCATCCTCATCGCCCAGGTGTTC 701

Qy       176 GlyTrpValLeuGlu------------------------TrpArgTrpLeuAlaValLeu 187
             |||      |||:::                        |||      |||      :::
Db       702 GGC------CTGGACTCCATCATGGGCAACAAGGACCTGTGGCCCCTGCTGCTGAGCATC 755

Qy       188 GlyCysValProProSerLeuMetLeuLeuLeuMetCysPheMetProGluThrProArg 207
                   :::|||      |||      :::::::::   |||   ||||||:::||||||
Db       756 ATCTTCATCCCGGCCCTGCTGCAGTGCATCGTGCTGCCCTTCTGCCCCGAGAGTCCCCGC 815

Qy       208 PheLeuLeu---ThrGlnHisArgArgGlnGluAlaMetAlaAlaLeuArgPheLeuTrp 226
             |||||||||      ::::::            |||   :::   |||:::   |||
Db       816 TTCCTGCTCATCAACCGCAACGAGGAGAACCGGGCCAAGAGTGTGCTAAAGAAGCTGCGC 875

Qy       227 GlySerGluGlnGlyTrpGluAspProProIleGlyAlaGluGlnSerPheHis------ 244
             |||:::               |||               |||:::|||
Db       876 GGGACAGCTGACGTGACCCATGACCTGCAGGAGATGAAGGAAGAGAGTCGGCAGATGATG 935

Qy       245 ---------------------LeuAlaLeuLeuArgGlnProGlyIleTyrLysProPhe 257
                                  |||   |||   |||   |||         :::|||
Db       936 CGGGAGAAGAAGGTCACCATCCTGGAGCTGTTCCGCTCCCCCGCCTACCGCCAGCCCATC 995

Qy       258 IleIleGlyValSerLeuMetAlaPheGlnGlnLeuSerGlyValAsnAlaValMetPhe 277
             :::|||   |||   |||         |||||||||||||||:::|||||||||   :::
Db       996 CTCATCGCTGTGGTGCTGCAGCTGTCCCAGCAGCTGTCTGGCATCAACGCTGTCTTCTAT 1055

Qy       278 TyrAlaGluThrIlePheGluGluAlaLysPheLysAspSerSerLeuAlaSerValVal 297
             |||:::   :::|||||||||:::|||      :::            |||::::::
Db      1056 TACTCCACGAGCATCTTCGAGAAGGCGGGGGTGCAGCAGCCTGTGTATGCCACCATTGGC 1115

Qy       298 ValGlyValIleGlnValLeuPheThrAlaValAlaAlaLeuIleMetAspArgAlaGly 317
                |||::::::         ||||||   |||:::      :::::::::|||||||||
Db      1116 TCCGGTATCGTCAACACGGCCTTCACTGTCGTGTCGCTGTTTGTGGTGGAGCGAGCAGGC 1175

Qy       318 ArgArgLeuLeuLeuValLeuSerGlyValValMetValPheSerThrSerAlaPheGly 337
             ||||||   |||   ::::::
Db      1176 CGGCGGACCCTGCACCTCATA--------------------------------------- 1196

Qy       338 AlaTyrPheLysLeuThrGlnGlyGlyProGlyAsnSerSerHisValAlaIleSerAla 357
                                     |||            :::      ||||||
Db      1197 ------------------------GGCCTCGCTGGCATGGCGGGTTGTGCCATACTCATG 1232

Qy       358 ProValSerAlaGlnProValAspAlaSerValGlyLeuAlaTrpLeuAlaValGlySer 377
                ::::::         ::::::            :::::::::|||::::::   :::
Db      1233 ACCATCGCGCTAGCACTGCTGGAGCAGCTACCCTGGATGTCCTATCTGAGCATCGTGGCC 1292

Qy       378 MetCysLeuPheIleAlaGlyPheAlaValGlyTrpGlyProIleProTrpLeuLeuMet 397
             :::      |||:::|||   |||   ||||||   |||||||||||||||   ::::::
Db      1293 ATCTTTGGCTTTGTGGCCTTCTTTGAAGTGGGTCCTGGCCCCATCCCATGGTTCATCGTG 1352

Qy       398 SerGluIlePheProLeuHisValLysGlyValAlaThrGlyIleCysValLeuThrAsn 417
             :::|||:::|||            :::      |||      :::         :::|||
Db      1353 GCTGAACTCTTCAGCCAGGGTCCACGTCCAGCTGCCATTGCCGTTGCAGGCTTCTCCAAC 1412

Qy       418 TrpLeuMetAlaPheLeuValThrLysGluPheSerSerLeuMetGluValLeuArgPro 437
             |||         |||:::|||         |||      :::   ::::::      |||
Db      1413 TGGACCTCAAATTTCATTGTGGGCATGTGCTTCCAGTATGTGGAGCAACTGTGTGGTCCC 1472

Qy       438 TyrGlyAlaPheTrpLeuAlaSerAlaPheCysIlePheSerValLeu------------ 453
             |||                        |||   ||||||:::||||||
Db      1473 TAC---------------------GTCTTCATCATCTTCACTGTGCTCCTGGTTCTGTTC 1511

Qy       454 ------PheThrLeuPheCysValProGluThrLysGlyLysThrLeuGluGlnIleThr 471
                   ||||||   |||   ||||||||||||||||||:::|||   ::::::|||
Db      1512 TTCATCTTCACCTACTTCAAAGTTCCTGAGACTAAAGGCCGGACCTTCGATGAGATCGCT 1571

Qy       472 AlaHisPhe 474
             :::   |||
Db      1572 TCCGGCTTC 1580
```

```
RESULT 12
US-09-489-039A-4560
; Sequence 4560, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
;  APPLICANT:  Gary Breton et. al
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
;  TITLE OF INVENTION:  PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  2709.2004001
;  CURRENT APPLICATION NUMBER: US/09/489,039A
;  CURRENT FILING DATE:  2000-01-27
;  PRIOR APPLICATION NUMBER:  US 60/117,747
;  PRIOR FILING DATE:  1999-01-29
;  NUMBER OF SEQ ID NOS:  14342
; SEQ ID NO 4560
;   LENGTH: 1506
;   TYPE: DNA
;   ORGANISM: Klebsiella pneumoniae
US-09-489-039A-4560

Alignment Scores:
Pred. No.:              7.32e-43      Length:          1506
Score:                  507.50        Matches:         149
Percent Similarity:     46.67%        Conservative:    75
Best Local Similarity:  31.04%        Mismatches:      188
Query Match:            20.66%        Indels:          69
DB:                     4             Gaps:            10

US-09-886-954A-1 (1-477) x US-09-489-039A-4560 (1-1506)

Qy        26 ValPheLeuAlaAlaPheAlaAlaAlaLeuGlyProLeuSerPheGlyPhe--------- 42
             |||:::               ||||||   |||   |||   ||||||:::
Db       106 GTCTGGACGATTTGTCTGGTCGCCGCCTGCGGTGGTTTACTGTTTGGCTATGACTGGGTG 165

Qy        43 AlaLeuGlyTyrSerSerProAlaIleProSerLeuGlnArgAlaAlaProProAlaPro 62
                :::|||   :::   |||         :::                  |||
Db       166 GTGATTGGCGGCGCTAAGCCATTTTATGAAGCCTGGTTTTCAATTACGGACCCG------ 219

Qy        63 ArgLeuAspAspAlaAlaAlaSerTrpPheGlyAlaValValThrLeuGlyAlaAlaAla 82
                         |||   :::   |||      :::         ||||||
Db       220 ------------GCGCAGTCCGGCTGGGCGATGAGCTCAGCGCTGTTGGGCTGTATTTTC 267

Qy        83 GlyGlyValLeuGlyGlyTrpLeuValAspArgAlaGlyArgLysLeuSerLeuLeuLeu 102
             |||   ::::::   ||||||      |||:::   ||||||||||||   |||:::|||
Db       268 GGCGCATTAATTTCGGGATGGTGCGCAGACAAACTGGGGCGCAAGCTGCCATTAATTCTT 327

Qy       103 CysSerValProPheValAlaGlyPheAlaValIleThrAlaAlaGlnAspValTrpMet 122
                :::|||   |||   |||                     |||            |||
Db       328 TCCGCCGTGCTGTTCAGCGCGTCGGCCTGGGGGACGGCGGTCGCCAGTCATTTCGATATG 387

Qy       123 LeuLeuGlyGlyArgLeuLeuThrGlyLeuAlaCysGlyValAlaSerLeuValAlaPro 142
                :::      |||::::::   |||:::      |||:::||||||   ::::::|||
Db       388 TTTGTGGTTTACCGCATTGTGGGCGGCGTAGGGATTGGTCTGGCTTCCGCTCTCAGCCCG 447

Qy       143 ValTyrIleSerGluIleAlaTyrProAlaValArgGlyLeuLeuGlySerCysValGln 162
             :::||||||:::|||::::::            ||||||         :::      |||
Db       448 CTTTACATTGCCGAAGTCAGCCCGGCAGAGAAAAGAGGACGTTTTGTCGCGGTCAATCAG 507

Qy       163 LeuMetValValValGlyIleLeuLeuAlaTyrLeuAlaGlyTrpValLeuGlu------ 180
             |||   :::|||:::|||:::|||   |||   |||         ::::::
Db       508 CTCACCATCGTGATTGGCGTGCTGGCCGCTCAGTTAATCAATCTGATGATTGCTGAACCG 567

Qy       181 ---------------------------------------------------------Trp 181
                                                                      |||
Db       568 GTGGAGCCGGGGGCGACGCAGCAGATGATTGTGGACAGCTGGAATGGGCAGATGGGCTGG 627

Qy       182 ArgTrpLeuAlaValLeuGlyCysValProProSerLeuMetLeuLeuLeuMetCysPhe 201
             ||||||:::               ||||||            |||:::||||||   |||
Db       628 CGCTGGATGTTCGGTGCGGAACTGGTGCCGGCACTGGCGTTTCTGGTCCTGATGTTTTTT 687

Qy       202 MetProGluThrProArgPheLeuLeuThrGlnHisArgArgGlnGluAlaMetAlaAla 221
```

```
              :::||||||:::||||||:::|||:::         :::   :::   |||   ||||||
Db        688 GTCCCCGAGTCGCCGCGCTGGCTCATGAAGGCCGGTAAACCGGAGCGCGCCCGGGCTGCG 747

Qy        222 LeuArgPheLeuTrpGlySerGluGlnGlyTrpGluAspPro------------------ 235
              |||      :::      ::::::         :::||||||
Db        748 CTGGAACGCATTGGTTCTGCCGACTATGCCGA-CAGGATCCTGCGTGAAATCGCGCATAC 806

Qy        236 ProIleGlyAlaGluGln-SerPheHisLeuAlaLeuLeuArgGlnProGlyIleTyrLy 255
              |||            |||                   |||      |||   :::
Db        807 CCTGGAAAAGGATAACAATAAAGTCTCCTACGGCGCGCTGTTGGCTCCCCAGGTGAAACC 866

Qy        255 sProPheIleIleGlyValSerLeuMetAlaPheGlnGlnLeuSerGlyValAsnAlaVa 275
                     |||||||||:::   |||      |||||||||      |||:::|||   ::
Db        867 GATTGTGATCATTGGCATGGTGCTCGCCATATTCCAGCAGTGGTGTGGGATTAACGTCAT 926

Qy        275 lMetPheTyrAlaGluThrIlePheGluGluAlaLysPheLys---AspSerSerLeuAl 294
              :      ||||||:::   ||||||      |||   |||      :::|||:::|||
Db        927 CTTTAACTACGCGCAGGAGATTTTTGCCTCGGCGGGGTTCGATATTAACAGCACGCTGAA 986

Qy        294 aSerValVal---ValGlyValIleGlnValLeuPheThrAlaValAlaAlaLeuIleMe 313
               |||:::|||      ||||||:::   ::::::||||||      |||      :::::
Db        987 ATCGATCGTCGCGACGGGCGTCGTTAATCTGGTCTTTACCATTGCGGCGCTGCCGCTGGT 1046

Qy        313 tAspArgAlaGlyArgArgLeuLeuLeuValLeu------SerGlyValValMetValPh 331
              :|||:::   |||||||||   |||::::::|||      ||||||:::   ::::::::
Db       1047 GGATAAAATCGGTCGCCGTAAATTAATGCTGCTTGGCGCTTCGGGATTGACGCTGATCTA 1106

Qy        331 eSerThrSerAlaPheGlyAlaTyrPheLysLeuThrGlnGlyGlyProGlyAsnSerSe 351
              :         |||   |||||||||               |||   |||
Db       1107 TGTGCTGATCGCC---GGCGCTTACGCCATGGGCATTATGGGGTGGCCG----------- 1152

Qy        351 rHisValAlaIleSerAlaProValSerAlaGlnProValAspAlaSerValGlyLeuAl 371
                                                               |||   |||
Db       1153 -------------------------------------------------GTACTGTTGCT 1163

Qy        371 aTrpLeuAlaValGlySerMetCysLeuPheIleAlaGlyPheAlaValGlyTrpGlyPr 391
                  ||||||                     ||||||   :::|||:::         ||
Db       1164 GGTGCTGGCGGCG------------------ATTGCTATTTATGCGTTGACCCTGGCGCC 1205

Qy        391 oIleProTrpLeuLeuMetSerGluIlePheProLeuHisValLysGlyValAlaThrGl 411
              |:::   |||:::|||::::::||||||||||||      |||:::|||:::|||
Db       1206 GGTGACCTGGGTGCTGCTGGCGGAGATTTTTCCCAACCGCGTACGTGGACTCGCGATGTC 1265

Qy        411 yIleCysValLeuThrAsnTrpLeuMetAlaPheLeuValThrLysGluPheSerSerLe 431
               :::      |||      |||:::      ||||||:::|||      |||      ||
Db       1266 TTTAGGTACCCTGGCGCTGTGGATCGCCTGTTTCCTGTTAACCTATACTTTCCCGCTGCT 1325

Qy        431 uMetGluValLeuArgProTyrGlyAlaPheTrpLeuAlaSerAlaPheCysIlePheSe 451
              |         |||         |||:::|||   |||            |||
Db       1326 TAATGCCGGTCTGGGCGCGGCAGGAAGCTTCCTGCTGTATGGTGTTATTTGCGCTGCCGG 1385

Qy        451 rValLeuPheThrLeuPheCysValProGluThrLysGlyLysThrLeuGluGlnIle 470
                  |||:::   |||      ||||||||||||||||||   |||||||||   :::
Db       1386 CTACCTCTATATCCTGCGCAACGTCCCGGAAACGAAAGGTATCACCCTCGAAGCGCTG 1443
```

RESULT 13
US-09-291-922-19
; Sequence 19, Application US/09291922
; Patent No. 6383776
; GENERAL INFORMATION:
; APPLICANT: Allen, Steve
; APPLICANT: Hitz, Bill
; APPLICANT: Kinney, Tony
; APPLICANT: Tingey, Scott
; TITLE OF INVENTION: Plant Sugar Transport Proteins
; FILE REFERENCE: BB-1163
; CURRENT APPLICATION NUMBER: US/09/291,922
; CURRENT FILING DATE: 1999-04-14
; EARLIER APPLICATION NUMBER: 60/083,044
; EARLIER FILING DATE: April 24, 1998
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 19
; LENGTH: 1914
; TYPE: DNA
; ORGANISM: Zea mays
US-09-291-922-19

```
Alignment Scores:
Pred. No.:                 1.35e-42        Length:          1914
Score:                     506.50          Matches:         137
Percent Similarity:        46.25%          Conservative:    91
Best Local Similarity:     27.79%          Mismatches:      182
Query Match:               20.61%          Indels:          83
DB:                        4               Gaps:            9

US-09-886-954A-1 (1-477) x US-09-291-922-19 (1-1914)

Qy         28 LeuAlaAlaPheAlaAlaAlaLeuGlyProLeuSerPheGlyPheAlaLeuGlyTyrSer 47
              :::   |||      |||::::::      :::      |||:::   :::|||
Db        122 ATATGTGCCATCCTGGCCTCCATGGCCTCTGTCATCCTTGGCTATGACATTGGGGTGATG 181

Qy         48 SerProAlaIleProSerLeuGlnArgAlaAlaProProAlaProArgLeuAspAspAla 67
              |||   |||         :::::::::                     :::   |||
Db        182 AGTGGAGCGGCCATGTACATCAAGAAGGACCTG------------AATATCACGGACGTG 229

Qy         68 AlaAlaSerTrpPheGlyAlaValValThrLeuGlyAlaAlaAlaGlyGlyValLeuGly 87
                                   :::::::::|||   :::      |||         |||
Db        230 CAGCTGGAGATCCTGATCGGGATCCTCAGTCTCTACTCGCTGTTCGGATCCTTCGCTGGC 289

Qy         88 GlyTrpLeuValAspArgAlaGlyArgLysLeuSerLeuLeuLeuCysSerValProPhe 107
                          ||||||   ||||||:::|||:::::::::      :::|||   |||
Db        290 GCGCGGACGTCCGACAGGATCGGGCGCCGCTTGACCGTCGTGTTCGCCGCTGTCATCTTC 349

Qy        108 ValAlaGlyPheAlaValIleThrAlaAlaGlnAspValTrpMetLeuLeuGlyGlyArg 127
                    |||      ::::::      |||   :::      ||||||:::   ||||||
Db        350 TTCGTGGGCTCGTTGCTCATGGGTTTCGCCGTCAACTACGGCATGCTCATGGCGGGCCGC 409

Qy        128 LeuLeuThrGlyLeuAlaCysGlyValAlaSerLeuValAlaProValTyrIleSerGlu 147
                 :::   |||:::      |||         ::::::||||||||||||   :::|||
Db        410 TTCGTGGCCGGAGTCGGTGTGGGCTACGGGGGCATGATCGCGCCCGTGTACACGGCCGAG 469

Qy        148 IleAlaTyrProAlaValArgGlyLeuLeuGlySerCysValGlnLeuMetValValVal 167
              |||:::      |||   ||||||   |||   :::      ::::::   :::   :::
Db        470 ATCTCGCCTGCGGCGTCCCGTGGCTTCCTGACCACCTTCCCGGAGGTGTTCATCAACATC 529

Qy        168 GlyIleLeuLeuAlaTyrLeuAlaGlyTrpVal---------------------LeuGlu 180
              ||||||||||||   ||||||:::   :::                        |||
Db        530 GGCATCCTGCTTGGCTACCTGTCCAACTTCGCGTTCGCGCGCCTCCCGCTCCACCTCGGC 589

Qy        181 TrpArgTrpLeuAlaValLeuGlyCysValProProSerLeuMetLeuLeuLeuMetCys 200
              ||||||   :::      :::|||   ||||||      |||:::   ||||||:::
Db        590 TGGCGCGTCATGCTCGCCATTGGCGCAGTTCCGTCCGGCCTGCTCGCGCTCCTGGTGTTC 649

Qy        201 PheMetProGluThrProArgPheLeuLeuThrGlnHisArgArgGlnGluAlaMetAla 220
                 |||||||||:::||||||:::|||:::   :::   |||      :::|||   |||
Db        650 TGCATGCCCGAGTCGCCTCGGTGGCTGGTCTTGAAGGGCCGCCTCGCGGACGCCAGGGCT 709

Qy        221 AlaLeuArg--------------------------------------------------- 223
                 |||
Db        710 GTGCTAGAGAAGACCTCTGCCACGCCAGAGGAGGCCGCCGAGCGGCTGGCCGACATCAAG 769

Qy        224 ------------------------------------------PheLeuTrpGlySerGlu 229
                                                           :::   |||   |||
Db        770 GCCGCGGCGGGGATTCCGAAGGGCCTCGACGGGGACGTAGTCACCGTACCCGGCAAGGAG 829

Qy        230 GlnGly------------------TrpGluAspProProIleGlyAlaGluGlnSerPhe 243
              ||||||                  |||:::
Db        830 CAAGGCGGCGGTGAGTTGCAGGTGTGGAAG------------------------------ 859

Qy        244 HisLeuAlaLeuLeuArgGlnProGlyIleTyrLysProPheIleIleGlyValSerLeu 263
                 |||   |||         |||   :::   :::      :::      |||   |||
Db        860 AAGCTCATCCTGTCCCCGACCCCGGCTGTCCGACGCATACTGCTCTCGGCCGTGGGTCTC 919

Qy        264 MetAlaPheGlnGlnLeuSerGlyValAsnAlaValMetPheTyrAlaGluThrIlePhe 283
                    |||||||||   ||||||   ::::::|||:::   |||:::      :::|||
```

```
Db         920 CACTTCTTCCAGCAGGCTTCTGGCAGCGACTCCGTCGTCCAGTACAGCGCCCGCCTGTTC 979

Qy         284 GluGluAlaLysPheLysAspSerSerLeu---------AlaSerValValValGlyVal 300
               :::   |||         |||   :::               :::      |||||||||
Db         980 AAGAGCGCGGGGATCACCGACGACAACAAGCTCCTGGGCGTCACCTGCGCGGTGGGCGTG 1039

Qy         301 IleGlnValLeuPheThrAlaValAlaAlaLeuIleMetAspArgAlaGlyArgArgLeu 320
                  :::      |||      ||||||      ::::::||||||||||||||||||
Db        1040 ACCAAGACGTTCTTCATCCTGGTGGCCACGTTCCTGCTGGACCGCGCGGGGCGTCGGCCT 1099

Qy         321 LeuLeuValLeuSerGlyValValMetValPheSerThrSerAlaPheGlyAlaTyrPhe 340
               ||||||::::::|||         |||:::   |||            |||:::
Db        1100 CTGCTGCTGATCAGCACGGGCGGGATGATTGTCTCGCTCATCTGCCTCGGGTCG------ 1153

Qy         341 LysLeuThrGlnGlyGlyProGlyAsnSerSerHisValAlaIleSerAlaProValSer 360
                                                                     |||:::
Db        1154 ---------------------------------------------GGGCTCACCGTCGCG 1168

Qy         361 AlaGlnProValAspAlaSerValGlyLeuAlaTrp---------LeuAlaValGlySer 377
                           |||      |||      ||||||         |||   :::   |||
Db        1169 GGGCATCACCCGGACACCAAGGTC------GCGTGGCCGTCGCCCTGTGCATCGCGTCA 1222

Qy         378 MetCysLeuPheIleAlaGlyPheAlaValGlyTrpGlyProIleProTrpLeuLeuMet 397
                        :::||||||   |||::::::|||   |||||||||      :::
Db        1223 ACCCTGTCCTACATCGCCTTCTTCTCCATCGGCCTCGGGCCCATCACGGGCGTGTACACC 1282

Qy         398 SerGluIlePheProLeuHisValLysGlyValAlaThrGlyIleCysValLeuThrAsn 417
               ||||||||||||||||||   |||:::   :::         :::   |||   :::|||
Db        1283 TCGGAAATATTCCCGCTGCAGGTGCGCGCGCTGGGCTTCGCGGTGGGTGTGGCGAGCAAC 1342

Qy         418 TrpLeuMetAlaPheLeuValThrLysGluPheSerSerLeuMetGluValLeuArgPro 437
                  :::   :::   :::::::::      |||   ||||||   :::   :::
Db        1343 CGCGTCACCAGCGCCGTCATCTCCATGACCTTCCTGTCCCTCTCCAAGGCCATCACCATC 1402

Qy         438 TyrGlyAlaPheTrpLeuAlaSerAlaPheCysIlePheSerValLeuPheThrLeuPhe 457
                  |||:::|||:::|||   |||               :::   :::|||
Db        1403 GGCGGCAGCTTCTTCCTCTACTCCGGCATCGCCGCGGTCGCTTGGGTTTTCTTCTTCACG 1462

Qy         458 CysValProGluThrLysGlyLysThrLeuGluGlnIle 470
               |||:::|||||||||:::|||:::|||||||||::::::
Db        1463 TGCCTCCCGGAGACACGCGGCCGGACGCTGGAGGAGATG 1501

RESULT 14
US-09-591-025-8
; Sequence 8, Application US/09591025
; Patent No. 6303373
; GENERAL INFORMATION:
;  APPLICANT:  Bogan, Jonathan S.
;  APPLICANT:  Lodish, Harvey F.
;  TITLE OF INVENTION:  Method of Measuring Plasma Membrane
;  TITLE OF INVENTION:  Targeting of GLUT4
;  FILE REFERENCE:  0399.1210-004
;  CURRENT APPLICATION NUMBER: US/09/591,025
;  CURRENT FILING DATE:  2000-06-09
;  PRIOR APPLICATION NUMBER:  60/154,078
;  PRIOR FILING DATE:  1999-09-15
;  PRIOR APPLICATION NUMBER:  60/138,237
;  PRIOR FILING DATE:  1999-06-09
;  NUMBER OF SEQ ID NOS:  8
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 8
;   LENGTH: 2592
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: modified GLUT4 containing myc tag sequences
US-09-591-025-8

Alignment Scores:
Pred. No.:              1.31e-41        Length:         2592
Score:                  499.00          Matches:        155
Percent Similarity:     48.07%          Conservative:   82
Best Local Similarity:  31.44%          Mismatches:     173
```

```
Query Match:            20.31%          Indels:         84
DB:                     4               Gaps:           15

US-09-886-954A-1 (1-477) x US-09-591-025-8 (1-2592)

Qy           9 ThrGlnProLeuLeuGlyProProGlyGlySerAlaProArgGlyArgArgValPheLeu 28
               ||||||   ||||||               |||      |||   :::   :::||||||
Db         455 ACTCAAGACCTCTTG---------------AGCCTGCAGAGGAGCAAAAGCTTATTTCTG 499

Qy          29 AlaAlaPheAlaAlaAlaLeuGlyProLeuSerPheGlyPheAlaLeuGlyTyrSerSer 48

Db         500 AAGAGGACTTGC----------------------------------------TTAAGGAC 520

Qy          49 ProAlaIleProSerLeuGlnArgAlaAlaProProAlaProArgLeuAspAspAlaAla 68
               ||||||   ||||||||||||      :::||||||:::      |||      |||
Db         521 CCAGCT---CCATCCCTCCAGGCACCCTCACCACCCTCTGG-GCCCTCTCCGTGGCCATC 576

Qy          69 AlaSerTrpPheGlyAlaValValThrLeuGlyAlaAlaAlaGlyGlyValLeuGlyGly 88
                  |||      |||   :::            :::         |||::::::
Db         577 TTTTCCGTGGCGGCATGATT---------TCCTCCTTCCTCATTGGTATCATCTCTCAG 627

Qy          89 TrpLeuValAspArgAlaGlyArgLysLeuSerLeuLeuLeuCysSerValProPheVal 108
               ||||||            |||||||||   ::::::|||:::   :::|||      |||
Db         628 TGGCTT------------GGAAGGAAAAGGGCCATGCTGGTCAACAATGTCCTGGCGGTG 675

Qy         109 AlaGly---------PheAlaValIleThrAlaAlaGlnAspValTrpMetLeuLeuGly 125
                  |||               :::      ||||||            ||||||:::
Db         676 CTGGGGGGCAGCCTCATGGGCCTGGCCAACGCTGCTGCCTCCTATGAAATGCTCATCCTT 735

Qy         126 GlyArgLeuLeuThrGlyLeuAlaCysGlyValAlaSerLeuValAlaProValTyrIle 145
               ||||||   |||   |||         |||:::   |||   :::   |||:::|||:::
Db         736 GGACGATTCCTCATTGGCGCCTACTCAGGGCTGACATCAGGGCTGGTGCCCATGTACGTG 795

Qy         146 SerGluIleAlaTyrProAlaValArgGlyLeuLeuGlySerCysValGlnLeuMetVal 165
                  |||||||||         :::||||||   ||||||:::      ||||||   :::
Db         796 GGGGAGATTGCTCCCACTCACCTGCGGGGCGCCCTGGGGACGCTCAACCAACTGGCCATT 855

Qy         166 ValValGlyIleLeuLeuAlaTyrLeuAlaGlyTrpValLeuGlu--------------- 180
               |||:::|||||||||:::|||   :::   |||      ||||||
Db         856 GTTATCGGCATTCTGATCGCCCAGGTGCTGGGC------TTGGAGTCCCTCCTGGGCACT 909

Qy         181 ---------TrpArgTrpLeuAlaValLeuGlyCysValProProSerLeuMetLeuLeu 197
                        |||      |||      |||      :::|||      |||   |||:::
Db         910 GCCAGCCTGTGGCCACTGCTCCTGGGCCTCACAGTGCTACCTGCCCTCCTGCAGCTGGTC 969

Qy         198 LeuMetCysPheMetProGluThrProArgPheLeu---LeuThrGlnHisArgArgGln 216
               |||:::   |||   ||||||:::||||||:::|||   :::   |||:::
Db         970 CTGCTGCCCTTCTGTCCCGAGAGCCCCCGCTACCTCTACATCATCCAGAATCTCGAGGGG 1029

Qy         217 GluAlaMetAlaAlaLeuArgPheLeu------TrpGlySerGluGlnGly--------- 231
                  |||      :::|||:::   |||      |||            |||
Db        1030 CCTGCCAGAAAGAGTCTGAAGCGCCTGACAGGCTGGCCGATGTTTCTGGAGTGCTGGCT 1089

Qy         232 ---TrpGluAspProProIleGlyAlaGluGlnSerPheHisLeuAlaLeuLeuArgGln 250
                     :::|||               |||:::         |||:::||||||:::
Db        1090 GAGCTGAAGGATGAGAAGCGGAAGCTGGAGCGTGAGCGGCCACTGTCCCTGCTCCAGCTC 1149

Qy         251 ProGlyIle---------TyrLysProPheIleIleGlyValSerLeuMetAlaPheGln 267
                  |||               :::|||   ||||||   |||   |||         |||
Db        1150 CTGGGCAGCCGTACCCACCGGCAGCCCCTGATCATTGCGGTCGTGCTGCAGCTGAGCCAG 1209

Qy         268 GlnLeuSerGlyValAsnAlaValMetPheTyrAlaGluThrIlePheGluGluAlaLys 287
               ||||||||||||:::|||||||||   :::|||:::   :::|||||||||   |||
Db        1210 CAGCTCTCTGGCATCAATGCTGTTTTCTATTATTCGACCAGCATCTTCGAGACAGCAGGG 1269

Qy         288 PheLysAspSerSerLeuAlaSerValValValGlyValIleGlnValLeuPheThrAla 307
                           :::   |||::::::      ||||||:::      :::||||||
Db        1270 GTAGGCCAGCCTGCCTATGCCACCATAGGAGCTGGTGTGGTCAACACAGTCTTCACCTTG 1329

Qy         308 ValAlaAlaLeuIleMetAspArgAlaGlyArgArgLeuLeuLeuValLeuSerGlyVal 327
               |||:::   |||:::::::::|||||||||||||||   |||   :::|||
Db        1330 GTCTCGGTGTTGTTGGTGGAGCGGGCGGGGCGCCGGACGCTCCATCTCCTG--------- 1380
```

```
Qy         328 ValMetValPheSerThrSerAlaPheGlyAlaTyrPheLysLeuThrGlnGlyGlyPro 347
                                                                     |||
Db        1381 ------------------------------------------------------GGCCTG 1386

Qy         348 GlyAsnSerSerHisValAlaIleSerAlaProValSerAlaGlnProValAspAlaSer 367
                                 ||||||         |||:::         ::::::
Db        1387 GCGGGCATGTGTGGCTGTGCCATCCTGATGACTGTGGCTCTGCTCCTGCTGGAGCGAGTT 1446

Qy         368 ValGlyLeuAlaTrpLeuAlaValGlySerMetCysLeuPheIleAlaGlyPheAlaVal 387
                     ::::::::::::::::::   ::::::      |||:::|||   |||   :::
Db        1447 CCAGCCATGAGCTACGTCTCCATTGTGGCCATCTTTGGCTTCGTGGCATTTTTGAGATT 1506

Qy         388 GlyTrpGlyProIleProTrpLeuLeuMetSerGluIlePheProLeuHisValLysGly 407
               |||   |||||||||||||||   ::::::::::|||:::|||            :::
Db        1507 GGCCCTGGCCCCATTCCTTGGTTCATCGTGGCCGAGCTCTTCAGCCAGGGACCCCGCCCG 1566

Qy         408 ValAlaThrGlyIleCysValLeuThrAsnTrpLeuMetAlaPheLeuValThrLysGlu 427
                  |||      :::         :::||||||         |||::::::
Db        1567 GCAGCCATGGCTGTGGCTGGTTTCTCCAACTGGACGAGCAACTTCATCATTGGCATGGGT 1626

Qy         428 PheSerSerLeuMetGluValLeuArgProTyrGlyAlaPheTrpLeuAlaSerAlaPhe 447
               |||      :::   |||   :::   ||||||                        |||
Db        1627 TTCCAGTATGTTGCGGAGGCTATGGGGCCCTAC---------------------GTCTTC 1665

Qy         448 CysIlePheSerValLeu------------------PheThrLeuPheCysValProGlu 461
                  :::|||:::||||||                  ||||||         |||||||||
Db        1666 CTTCTATTTGCGGTCCTCCTGCTGGGCTTCTTCATCTTCACCTTCTTAAGAGTACCTGAA 1725

Qy         462 ThrLysGlyLysThrLeuGluGlnIleThrAlaHisPhe 474
               |||:::|||:::|||   :::||||||:::|||   |||
Db        1726 ACTCGAGGCCGGACGTTTGACCAGATCTCGGCTGCCTTC 1764

RESULT 15
US-09-919-172-79
; Sequence 79, Application US/09919172
; Patent No. 6673545
; GENERAL INFORMATION:
;  APPLICANT: Faris, Mary
;  APPLICANT:  Turner, Christopher M.
;  TITLE OF INVENTION: PROSTATE CANCER MARKERS
;  FILE REFERENCE: PA-0036 US
;  CURRENT APPLICATION NUMBER: US/09/919,172
;  CURRENT FILING DATE:  2001-07-30
;  PRIOR APPLICATION NUMBER: 60/222,469
;  PRIOR FILING DATE: 2000-07-28
;  NUMBER OF SEQ ID NOS: 102
;  SOFTWARE: PERL Program
; SEQ ID NO 79
;   LENGTH: 5227
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. 6673545 410785.1
;   NAME/KEY: unsure
;   LOCATION: 4928, 4934, 4939, 4944, 4973, 4992
;   OTHER INFORMATION: a, t, c, g, or other
US-09-919-172-79

Alignment Scores:
Pred. No.:                5e-41          Length:         5227
Score:                    498.00         Matches:        147
Percent Similarity:       45.99%         Conservative:   94
Best Local Similarity:    28.05%         Mismatches:     183
Query Match:              20.27%         Indels:         100
DB:                       4              Gaps:           15

US-09-886-954A-1 (1-477) x US-09-919-172-79 (1-5227)

Qy          23 GlyArgArgValPheLeuAlaAlaPheAlaAlaAlaLeuGlyProLeuSerPheGlyPhe 42
               |||      ||||||               |||   ||||||         ||||||:::
Db          60 GGGACCCTGGTTTTC---ACTGTCATCACTGCTGTGCTGGGTTCCTTCCAGTTTGGATAT 116

Qy          43 AlaLeuGlyTyr------------------------------------------------ 46
                  :::|||
Db         117 GACATTGGTGTGATCAATGCACCTCAACAGGTAATAATATCTCACTATAGACATGTTTTG 176

Qy          47 ------------------------------------------------SerSerProAla 50
                                                               |||:::
Db         177 GGTGTTCCACTGGATGACCGAAAAGCTATCAACAACTATGTTATCAACAGTACAGATGAA 236

Qy          51 IleProSerLeuGlnArgAlaAlaProProAlaPro------------ArgLeuAspAsp 66
               :::|||::::::      :::      |||   |||                  :::
Db         237 CTGCCCACAATCTCATACTCAATGAACCCAAAACCAACCCCTTGGGCTGAGGAAGAGACT 296

Qy          67 AlaAlaAlaSer------------------TrpPheGlyAlaVal------ValThrLeu 78
                  ||||||:::                  |||      :::|||            :::
Db         297 GTGGCAGCTGCTCAACTAATCACCATGCTCTGGTCCCTGTCTGTATCCAGCTTTGCAGTT 356

Qy          79 GlyAlaAlaAlaGlyGlyValLeuGlyGlyTrpLeuValAspArgAlaGlyArgLysLeu 98
               |||                     ||||||||||||   |||      ||||||
Db         357 GGTGGAATGACTGCATCATTCTTTGGTGGGTGGCTTGGGGACACACTTGGAAGAATCAAA 416

Qy          99 SerLeuLeuLeuCysSerVal---------------ProPheValAlaGlyPheAlaVal 113
               ::::::|||:::   ::::::                     :::   ||||||:::
Db         417 GCCATGTTAGTAGCAAACATTCTGTCATTAGTTGGAGCTCTCTTGATGGGGTTTCAAAA 476

Qy         114 IleThrAlaAlaGlnAspValTrpMetLeuLeuGlyGlyArgLeuLeuThrGlyLeuAla 133
               :::      :::            :::::::::   ||||||   ::::::||||||
Db         477 TTGGGACCATCTCAT------ATACTTATAATTGCTGGAAGAAGCATATCAGGACTATAT 530

Qy         134 CysGlyValAlaSerLeuValAlaProValTyrIleSerGluIleAlaTyrProAlaVal 153
               ||||||:::   |||   :::   |||:::||||||   |||||||||      |||:::
Db         531 TGTGGGCTAATTTCAGGCCTGGTTCCTATGTATATCGGTGAAATTGCTCCAACCGCTCTC 590

Qy         154 ArgGlyLeuLeuGlySerCysValGlnLeuMetValValValGlyIleLeuLeuAlaTyr 173
               ||||||   ||||||:::      ||||||   :::|||   |||||||||::::::
Db         591 AGGGGAGCACTTGGCACTTTTCATCAGCTGGCCATCGTCACGGGCATTCTTATTAGTCAG 650

Qy         174 LeuAlaGlyTrpValLeuGlu------------------------TrpArgTrpLeuAla 185
               :::   |||      ||||||                        |||      |||
Db         651 ATTATTGGT------CTTGAATTTATCTTGGGCAATTATGATCTGTGGCACATCTTGCTT 704

Qy         186 ValLeuGlyCysValProProSerLeuMetLeuLeuLeuMetCysPheMetProGluThr 205
                  |||      |||         |||      ||||||:::   |||   ||||||:::
Db         705 GGCCTGTCTGGTGTGCGAGCCATCCTTCAGTCTCTGCTACTCTTTTTCTGTCCAGAAAGC 764

Qy         206 ProArgPheLeu---LeuThrGlnHisArgArgGlnGluAlaMetAlaAlaLeuArgPhe 224
               ||||||:::|||   :::                  :::|||      :::|||:::
Db         765 CCCAGATACCTTTACATCAAGTTAGATGAGGAAGTCAAAGCAAAACAAAGCTTGAAAAGA 824

Qy         225 LeuTrpGly------------------------SerGluGlnGlyTrpGluAspProPro 236
               |||   |||                        :::|||         |||
Db         825 CTCAGAGGATATGATGATGTCACCAAAGATATTAATGAAATGAGAAAAGAAAGAGAAGAA 884

Qy         237 IleGlyAlaGluGlnSerPheHisLeuAlaLeuLeuArgGlnProGlyIleTyrLys--- 255
                     :::||||||         :::      |||               |||:::
Db         885 GCATCGAGTGAGCAGAAAGTCTCTATAATTCAGCTCTTCACCAATTCCAGCTACCGACAG 944

Qy         256 ProPheIleIleGlyValSerLeuMetAlaPheGlnGlnLeuSerGlyValAsnAlaVal 275
               |||   ::::::   :::   |||         ||||||   ||||||:::|||   :::
Db         945 CCTATTCTAGTGGCACTGATGCTGCATGTGGCTCAGCAATTTTCCGGAATCAATGGCATT 1004

Qy         276 MetPheTyrAlaGluThrIlePheGluGluAlaLysPheLysAspSerSerLeuAlaSer 295
                  :::|||:::   :::||||||:::   |||                     |||:::
Db        1005 TTTTACTACTCAACCAGCATTTTTCAGACGGCTGGTATCAGCAAACCTGTTTATGCAACC 1064

Qy         296 ValValValGlyValIleGlnValLeuPheThrAlaValAlaAlaLeuIleMetAspArg 315
               :::   ||||||   :::   ::::::||||||||||||:::      ::::::::::::
Db        1065 ATTGGAGTTGGCGCTGTAAACATGGTTTTCACTGCTGTCTCTGTATTCCTTGTGGAGAAG 1124

Qy         316 AlaGlyArgArg------LeuLeuLeuValLeuSerGlyValValMetValPheSerThr 333
               ||||||||||||         |||:::   :::||||||:::   :::
Db        1125 GCAGGGCGACGTTCTCTCTTTCTAATTGGAATGAGTGGGATGTTTGTT------------ 1172

Qy         334 SerAlaPheGlyAlaTyrPheLysLeuThrGlnGlyGlyProGlyAsnSerSerHisVal 353
```

```
Db        1173 ---------------------------------------------------------TGT 1175

Qy         354 AlaIleSerAlaProValSerAlaGlnProValAsp-AlaSerValGlyLeuAlaTrp-L 373
               ||||||         |||            :::::: :::|||:::||||||:::::: :
Db        1176 GCCATCTTCATGTCAGTGGGACTTGTGCTGCTGAATAAGTTCTCTTGGACTGAGTTACTG 1235

Qy         373 euAlaValGlySerMetCysLeuPheIleAlaGlyPheAlaValGlyTrpGlyProIleP 393
               ::::::::   ::::::   ||||||::::::   |||   :::|||   ||||||||||
Db        1236 TGAGCATGATAGCCATCTTCCTCTTTGTCAGCTTCTTTGAAATTGGGCCAGGCCCGATCC 1295

Qy         393 roTrpLeuLeuMetSerGluIlePheProLeuHisValLysGlyValAlaThrGlyIleC 413
               |||||   :::::::::|||   |||            :::      |||      |||
Db        1296 CCTGGTTCATGGTGGCTGAGTTTTTCAGTCAAGGACCACGTCCTGCTGCTTTAGCAATAG 1355

Qy         413 ysValLeuThrAsnTrpLeuMetAlaPheLeuValThrLysGluPheSerSerLeuMetG 433
                       :::||||||         |||:::|||         |||      :::   :
Db        1356 CTGCATTCAGCAATTGGACCTGCAATTTCATTGTAGCTCTGTGTTTCCAGTACATTGCGG 1415

Qy         433 luValLeuArgProTyrGlyAlaPheTrpLeuAlaSerAlaPheCysIlePheSerValL 453
               ::         ||||||      |||      |||               |||:::   |
Db        1416 ACTTCTGTGGACCTTATGTGTTTTTCCTCTTTGCTGGAGTGCTCCTGGCCTTTACC---C 1472

Qy         453 euPheThrLeuPheCysValProGluThrLysGlyLysThrLeuGluGlnIleThrAlaH 473
               ||||||||   |||   |||||||||||||||||||||:::   |||:::|||   |||
Db        1473 TGTTTACATTTTTTAAAGTTCCAGAAACCAAAGGAAAGTCTTTTGAGGAAATTGCTGCAG 1532

Qy         473 isPheGlu 475
                 |||:::
Db        1533 AATTCCAA 1540

Search completed: September 28, 2004, 04:18:30
Job time : 152 secs
```